JMB

Disallowed Ramachandran Conformations of Amino Acid Residues in Protein Structures

K. Gunasekaran, C. Ramakrishnan and P. Balaram*

*Molecular Biophysics Unit
Indian Institute of Science
Bangalore, India*

An analysis of the nature and distribution of disallowed Ramachandran conformations of amino acid residues observed in high resolution protein crystal structures has been carried out. A data set consisting of 110 high resolution, non-homologous, protein crystal structures from the Brookhaven Protein Data Bank was examined. The data set consisted of a total of 18,708 non-Gly residues, which were characterized on the basis of their backbone dihedral angles (ϕ, ψ). Residues falling outside the defined "broad allowed limits" on the Ramachandran map were chosen and the reported B-factor value of the α-carbon atom was used to further select well defined disallowed conformations. The conformations of the selected 66 disallowed residues clustered in distinct regions of the Ramachandran map indicating that specific ϕ, ψ angle distortions are preferred under compulsions imposed by local constraints. The distribution of various amino acid residues in the disallowed residue data set showed a predominance of small polar/charged residues, with bulky hydrophobic residues being infrequent. As a further check, for all the 66 cases non-hydrogen van der Waals short contacts in the protein structures were evaluated and compared with the ideal "Ala-dipeptide" constructed using disallowed dihedral angle (ϕ, ψ) values. The analysis reveals that short contacts are eliminated in most cases by local distortions of bond angles. An analysis of the conformation of the identified disallowed residues in related protein structures reveals instances of conservation of unusual stereochemistry.

*Corresponding author

Introduction

The Ramachandran map (Ramachandran *et al.*, 1963; Ramakrishnan & Ramachandran, 1965; Ramachandran & Sasisekharan, 1968) provides a convenient method for the analysis of the stereochemistry of the polypeptide chain backbone in protein structures. Conformational analysis in dihedral (ϕ, ψ) angle space is particularly powerful for large structures where each individual residue is structurally characterized by a single set of ϕ, ψ values. The range of accessible ϕ, ψ values at each non-Gly amino acid residue in a protein structure is sharply restricted by local unfavorable steric interactions. In experimentally determined protein structures the vast majority of non-Gly residues adopt ϕ, ψ values that lie within the allowed

regions of the Ramachandran map for an idealized "L-Ala-dipeptide" (Herzberg & Moult, 1991; Weaver *et al.*, 1990; MacArthur *et al.*, 1994; Morris *et al.*, 1992; Thornton, 1992). Indeed, the stereochemical quality of a protein model may be judged by the use of ϕ, ψ scatter plots, with incorrect structures generally having a much larger fraction of residues lying in disallowed regions. An interesting feature that has emerged from a large number of high resolution protein structure determinations is that specific residues can sometimes adopt disallowed ϕ, ψ values, suggesting that unfavorable local interactions may be offset by other compensating factors (Stites *et al.*, 1994; Jia *et al.*, 1993). A recent study has attempted to quantify the strain energy associated with unfavorable ϕ, ψ values by analyzing the stability of suitably chosen Gly/Ala mutants in staphylococcal nuclease (Stites *et al.*, 1994). The availability of a large data set consisting of well refined high
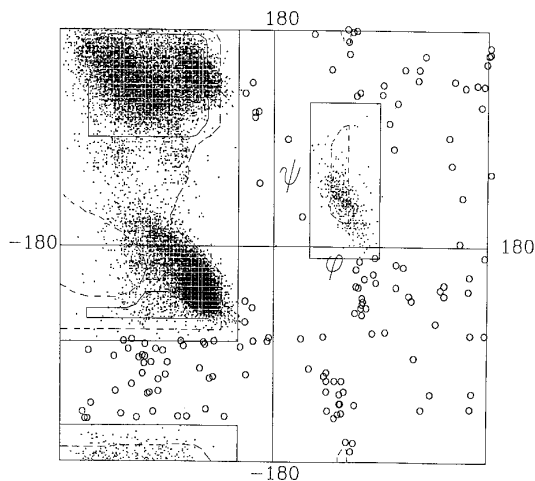

Figure 1. Distribution of observed dihedral angles ϕ, $\psi^{(o)}$ of non-Gly residues (18,708) in the 110, high resolution, largely non-homologous, protein structures selected from the Brookhaven Protein Data Bank. A definition of broad "allowed regions" are shown as rectangles, residues which lie outside these regions are shown as (o); allowed residues are shown as (·). For details see Methods.

resolution protein crystal structures prompted us to examine the distribution and nature of disallowed Ramachandran residues. The results described in this paper establish that only a very small fraction of residues (less than 0.8%) lie in appreciably disallowed regions with small polar residues having a significantly greater probability of adopting unusual backbone conformations.

Results and Discussion

Sterically disallowed clusters

From the data set of 110 largely non-homologous protein crystal structures a total of 18,708 non-Gly residues yielded 146 residues lying in the disallowed regions, as indicated in Figure 1. It should be noted that the allowed region shown in Figure 1 has slightly larger areas of ϕ, ψ space (39.5%) than the region defined using the extreme limit contact distances (25%; Ramachandran *et al.*, 1963). This relaxation was used so as to restrict the analysis to residues having significant distortions in ϕ, ψ space. Figure 2 provides the summary of the procedure used to extract the 66 disallowed residues. Figure 3 summarizes the distribution of the 66 disallowed Ramachandran residues. It is evident that clusters can be found in specific regions of disallowed ϕ, ψ space (Figure 3). The disallowed residues, with the exception of only three residues, can be clustered in six regions of which the cluster VI forms a minor region that is actually allowed by the Ramachandran extreme limit contact distance criteria. These residues are not considered further. Table 1 lists the disallowed residues occurring in different clusters. Clusters II and III in the lower right hand quadrant (positive
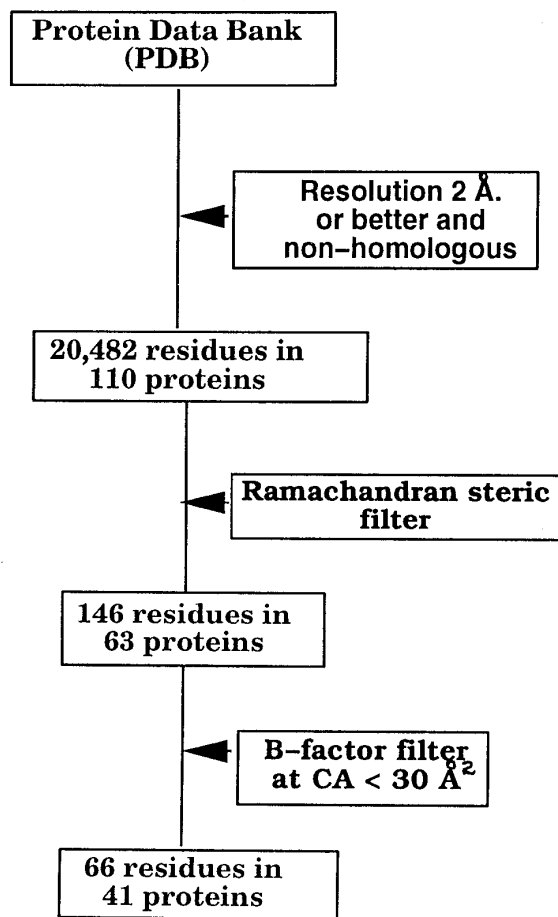

Figure 2. Flow chart depicting the scheme used in the selection of 66 disallowed residues from the Brookhaven Protein Data Bank. For details see Methods.

ϕ, negative ψ) correspond to regions which are populated by residues adopting the $i + 1$ positions of type II' β-turn conformations: $\phi = 60°$, $\psi = -120°$
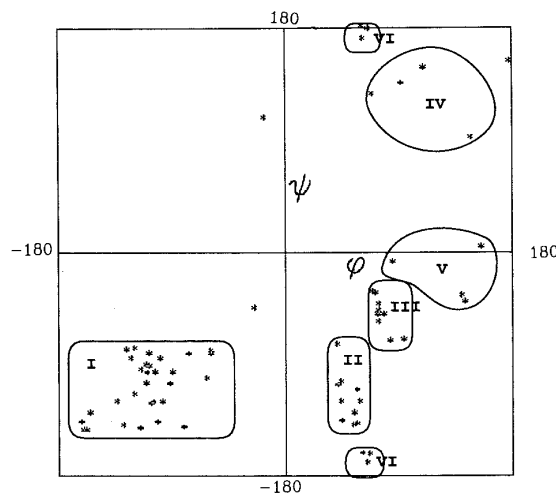

Figure 3. Plot of backbone dihedral angles ϕ, ψ (°) of the disallowed residues selected using the scheme shown in Figure 2 with clusters being marked and identified with Roman numerals.

Table 1. List of disallowed Ramachandran residues

Protein code	Residue number	Residue	Torsion angle ϕ, ψ (°)	Protein code	Residue number	Residue	Torsion angle ϕ, ψ (°)
Cluster I				2fb4 L	153	Asp	(54, −115)
1ccr	35	Lys	(−132, −143)	3bcl	327	Asx	(56, −124)
1cse E	32	Asp	(−161, −147)	3blm	69	Ala	(42, −140)
1fkf	81	Ala	(−137, −124)	3ebx	8	Ser	(41, −124)
1gox	258	Gln	(−93, −100)	3tln	159	Asn	(54, −142)
1gox	78	Thr	(−114, −109)	5tim A	13	Lys	(51, −143)
1sn3	33	Asn	(−62, −85)	Cluster III			
1ton	35	Asn	(−158, −134)	1rbp	111	Tyr	(66, −36)
2alp	39	Ala	(−130, −82)	2fb4 L	50	Asp	(70, −54)
2cdv	51	Cys	(−101, −124)	2fbj L	50	Ile	(71, −51)
2csc	67	Arg	(−80, −86)	2gbp	91	Asn	(68, −37)
2csc	274	His	(−110, −126)	2rhe	52	Asn	(71, −46)
2fb4 L	27	Asn	(−112, −85)	2sga	100	Asn	(80, −75)
2pka B	214	Ser	(−107, −100)	3c2c	92	Thr	(91, −74)
2prk	39	Asp	(−165, −141)	3tln	26	Thr	(75, −54)
2rhe	28	Asp	(−114, −94)	8apb	89	Asp	(71, −60)
2tec E	38	Asp	(−163, −147)	Cluster IV			
351c	22	Met	(−113, −101)	1ak3 A	198	Glu	(66, 123)
3blm	220	Leu	(−108, −125)	1gd1 O	237	Val	(89, 132)
3c2c	91	Met	(−63, −83)	1gp1 A	24	Glu	(144, 88)
3cox	231	Thr	(−117, −96)	2ltn A	27	Thr	(107, 145)
3est	171	Tyr	(−95, −110)	Cluster V			
3grs	52	His	(−123, −118)	3bcl	70	Ala	(82, −12)
3grs	219	His	(−119, −146)	5cpa	199	Ser	(153, 0)
3tln	152	Thr	(−112, −95)	2gbp	236	Asp	(140, −44)
4enl	320	Asp	(−123, −81)	8abp	232	Asn	(137, −39)
4ptp	71	Asp	(−126, −89)	Cluster VI			
5cpa	247	Ile	(−103, −89)	3tln	92	Ser	(62, −174)
5cpa	57	Ser	(−66, −105)	3tln	107	Ser	(58, −166)
7rsa	60	Gln	(−105, −141)	1gd1 O	147	Ala	(64, −167)
8abp	254	Pro	(−84, −145)	3grs	425	Asn	(63, 176)
Cluster II				3bcl	119	Ser	(59, 168)
1aap A	2	Arg	(38, −78)	2csc	239	Glu	(58, 178)
1gox	30	Glu	(39, −111)	Others			
1snc	138	Asn	(41, −108)	1gd1 O	186	Asp	(−20, 105)
2ca2	252	Lys	(52, −134)	1ova A	278	Ser	(−28, −49)
				1gox	79	Ala	(175, 150)

See Figure 3 for cluster identification.

(Venkatachalam, 1968; Wilmot & Thornton, 1988, 1990; Mattos *et al.*, 1994) and the γ-turn conformation: $\phi = 80°$, $\psi = -70°$ (Rose *et al.*, 1985; Milner-White, 1990). In both these conformations the formation of local backbone hydrogen bonds, $4 \rightarrow 1$ type for β-turn and $3 \rightarrow 1$ type for γ-turn, compensate the unfavorable van der Waals contact. An examination of the ten residues in cluster II (Figure 3) reveals that six residues were in fact at the $i + 2$ position of nearly ideal type II' β-turns. In one case (5TIM LysA13) the residue forms part of a chain interface with a significant change in the ψ_3 value of the type II' β-turn. Only very few residues were found in cluster IV (four residues) and cluster V (four residues) compared to cluster I (30 residues). Cluster V may be derived from highly distorted left handed helical (α_L) conformations.

Interestingly, clustering of disallowed residues in specific regions of ϕ, ψ space was much less evident before applying a *B*-factor cutoff for selecting residues with unfavorable backbone conformations. This can be seen from the distribution of residues lying outside the allowed regions in the Ramachandran map shown in Figure 1, which is representative of the total protein data set without application of any selection criteria. Clustering suggests that specific ϕ, ψ distortions are preferred under compulsions imposed by local constraints.

Amino acid propensities in disallowed regions

A histogram depicting the distribution of various residues in the disallowed regions is shown in Figure 4. Propensity values were computed using the formula:

$$P_i = \frac{\left(\dfrac{D_i}{\displaystyle\sum_{j=1}^{19} D_j} \right)}{\left(\dfrac{F_i}{\displaystyle\sum_{j=1}^{19} f_j} \right)}$$

where, D_i is the number of disallowed conformational points and F_i the total number of conformational points of residue i. A value >1 indicates a significant tendency to adopt a disallowed conformation, while values <1 suggest that
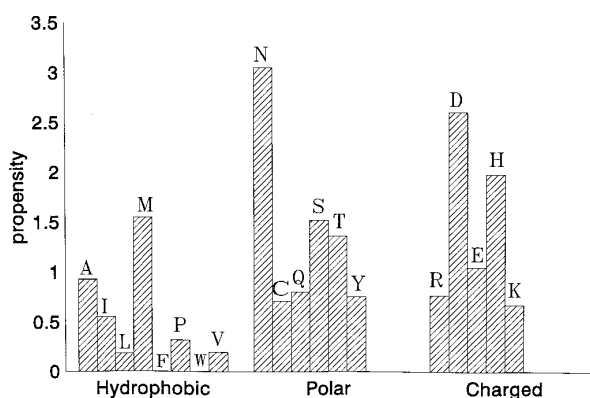
Amino acid distribution

Figure 4. Propensity distribution of non-Gly amino acids in 66 disallowed residues categorized as hydrophobic, polar and charged residues.

backbone distortions are extremely unlikely for these residues. Although the data set of disallowed residues is too small to permit statistically rigorous conclusions, some biochemically relevant features are readily apparent. Disallowed conformations are most often found for small polar/charged residues with the largest number of examples being observed for Asn, Asp, Ser, Thr and His residues. Bulky hydrophobic residues which are found predominantly in well packed interiors of proteins like Leu, Ile, Val and Phe do not generally adopt disallowed conformations. The bulky charged residues Glu, Lys and Arg also have a relatively low propensity for backbone distortions. An unusually high propensity for distortion was observed for Met; this may be an artifact due to an insufficient number of examples of Met in the total data set. The tendency of small polar residues to accommodate backbone distortions may be rationalized by their ability to form compensating hydrogen bond interactions involving functional side-chains. In the case of hydrophobic residues, compensating hydrogen bond interactions are restricted only to backbone amide groups. Indeed, inspection of the 66 disallowed residues revealed that, except in one case, main-chain–main-chain/main-chain–side-chain/side-chain–side-chain hydrogen bonds were formed (using the criteria for N– – –O and O– – –O distance of 2.4 to 3.6 Å). The only exception was in the case of carbonoxypeptidase (5CPA) Ser57, in which the side-chain had a relatively high *B*-factor value. Since in protein structures most hydrogen bonding groups are involved in intramolecular interactions or in making contacts with water molecules, it is difficult to unequivocally suggest that hydrogen bonding interactions do indeed offset the energy penalty that has to be paid for backbone stereochemistry distortions.

It was of interest to examine if Gly residues populate the "L-disallowed" region; out of a total of 1774 Gly residues in the data set, 638 Gly

residues were found in the L-disallowed regions yielding a propensity value of 10.34. An examination of a specific region which lies in the lower left region of the Ramachandran map ($\phi = -180°$ to $-30°$, $\psi = -150°$ to $-80°$) revealed that only 30 Gly residues adopted ϕ, ψ values in this region, resulting in a propensity value of 4.8 (calculated for this region). This region accommodates almost half (30 residues) the total L-residues found in the disallowed region (66 residues). Considering the population distribution of the amino acid residues in ϕ, ψ space, it appears that both L and Gly residues are only infrequently found in the cluster I regions.

van der Waals short contact check

As a further check, for all the 66 cases an "Ala-dipeptide" model was constructed using the ideal bond angle and bond length and the observed disallowed ϕ, ψ values in protein structure. Stereochemical non-hydrogen contacts were evaluated for all the ideal models and for the disallowed residues in the protein structures (Ramachandran *et al.*, 1963). In 42 cases no short van der Waals non-hydrogen contacts were observed though such short contacts existed in the ideal model. On closer examination, it was found that in all these cases the bond angles at C^α ($C^\beta - C^\alpha - C$ and $N - C^\alpha - C^\beta$) were significantly distorted from the ideal value. It is pertinent to note that the original Ramachandran map was computed using an ideal bond angle of $110°$ at C^α (Ramachandran *et al.*, 1963). The average value for the two bond angles in small molecule structures were $109°$ for $C^\beta - C^\alpha - C$ and $110°$ for $N - C^\alpha - C^\beta$ (Engh & Huber, 1991). The most frequent unfavorable contact involves C^β and C of the previous residue (four such short contacts were observed in the protein structures against 20 in the ideal models) or N of the next residue (18 contacts in the protein structures against 43 in the ideal models). It is thus clear that in many cases unfavorable van der Waals contacts are avoided by the bond angle distortions.

In 24 cases where short contacts existed in both protein structure and the ideal model, interestingly, not all the short contacts shown by the ideal model were present in the protein model. This further reinforced the view that unfavorable van der Waals contacts are often relieved by distortion of local geometry. In only four cases (1GOX Ala79, 2CSC Arg67, 3BCL Ser119 and 1SN3 Asn33) short contacts which were not present in the ideal model were found in the protein model, which was an unusual situation.

Location of disallowed residues

A significant number of the disallowed residues (25) occur in long irregular segments. Only 17 disallowed residues were involved in short segments ($\leqslant$ five residues) linking the secondary

structures; helix-helix (four), helix-strand (two), strand-helix (six) and strand-strand (five). Further, 12 residues were involved in terminating the helices (three) or strands (nine) and 12 residues were at the N-terminal end of helices (five) or strands (seven). Secondary structure identification was made on the basis of backbone ϕ, ψ values (Ramakrishnan & Srinivasan, 1990). Accessible surface areas (ASA) of the disallowed residues were computed (Lee & Richards, 1971) and these values were compared with the average ASA value obtained for each residue from the data base. The computed distribution of ASA values for disallowed residues showed that there is no overwhelming preference for high or low solvent accessibility except for the residue Asp for which nine out of 11 residues had ASA values less than 10 Å^2, as compared to the average value of 42 Å^2.

In the case of the enzymes in the protein data set which yielded the 66 disallowed residues, no clearly apparent relation between the active site and the unusual stereochemistry could be observed. However, the present disallowed residue data set provides examples which are at or near the active site, as defined in the original description of the structure. These include 1CSE AspE32, 2PRK Asp39, 2TEC AspE38, 4ENL Asp320, 4PTP Asp71 and 5TIM LysA13, which provide examples of a disallowed active site residue, while 1GD1 AlaO147, 1GOX Gln258, 2SGA Asn100 and 3BLM Ala69 are near the active site of the respective enzymes. It should be noted that unusual Ramachandran angles have been correlated with biochemical function in the case of the histidine-containing phosphocarrier protein (Jia *et al.*, 1993). In this case, Ala16 adopts a conformation in cluster VI (Figure 3) but this "high energy" state is compensated by three favorable hydrogen bonding interactions. During phosphorylation of His15, Ala16 is postulated to move to the allowed, right-handed helical (α_R), region of the conformational space.

Conservation of disallowed residues

Is the unusual stereochemistry observed at the disallowed residues an intrinsic feature dictated by the structural environment or is it a lingering artifact of the structure determination process, despite a rigorous attempt to screen the data base? (Kleywegt & Jones, 1995). The availability of structures of the same protein determined by different groups of investigators or of minor variants of a protein (complex with small molecules, mutant proteins) permits this question to be addressed. In the present study, the 66 disallowed residues occur in 41 distinct protein structures. We examined the protein data base for examples of structures which were directly related to these examples. Figure 5 summarizes the procedure by which the conformations of the 66 disallowed
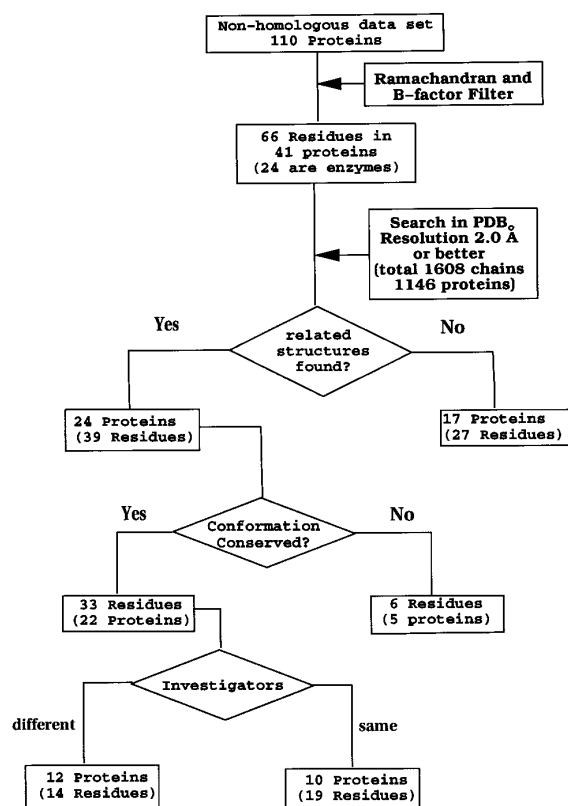

Figure 5. Flow chart depicting the procedure by which the conformations of the 66 disallowed residues in related structures are compared.

residues in different structures were compared. Cases where one structure was used in a molecular replacement solution of the other were not excluded. Of 41 proteins in the disallowed residue data set, related examples were found only for 24 cases, with a total of 39 disallowed residues available for comparison. Of these, 33 residues did have the disallowed conformation conserved, while six residues adopted allowed ϕ, ψ values in some of the related protein structures. On closer examination it was found that in five of these cases ψ values lie on the borderline between the allowed and disallowed regions of the Ramachandran map, with small variations resulting in a change in the status of the residue as defined by our criteria. Only three residues, Lys252 in carbonic anhydrase (2CA2, Figure 6(a), 6), Arg67 in citrate synthase (2CSC, Figure 6(b), 3) and Ser57 in carboxypeptidase (5CPA, Figure 6(c), 5) appear to have distinctly different ϕ, ψ values in the various structures. Of the 33 carbonic anhydrase structures, Lys252 has a ψ value between $-128°$ and $-141°$ in 26 structures and in seven structures between $51°$ and $60°$. Of the eight citrate synthase structures, Arg67 has α_R conformation in six structures and α_L conformation in two structures. Of the nine carboxypeptidase structures, Ser57 has a ψ value between $-87°$ and $-118°$ in five structures and in four structures between $119°$ and $124°$. It is clear that the conformations correspond to approximately $180°$
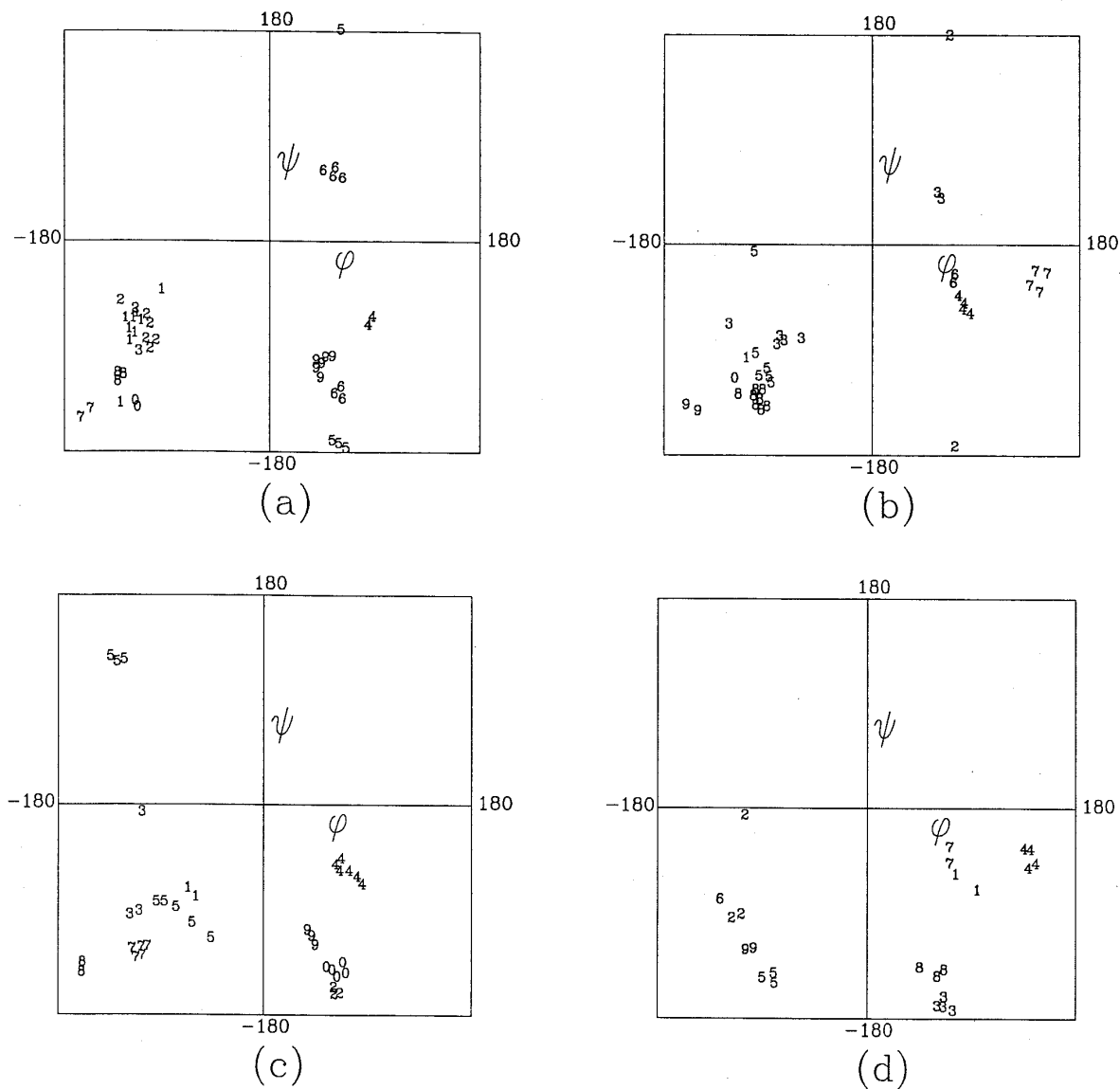
Figure 6. Conservation of the unusual residue conformations in homologous proteins and variants. Numbers represents protein families. Protein codes, which are present in the original 110 proteins data set, are indicated in bold letters (below). In cases where a large number of available examples cluster at the same positions, only a few are marked for the sake of clarity. (a) 1, beta trypsin (Asp71) **4PTP** 1TGN 1NTP 1PPC 1TGB 1TGC 1TGT 1TLD 1TPA 1TPO 1TPP 2PTN 2TGA 2TGP 2TGT 3PTN 3TPI 1TPP 1GBT 3PTB 1TLD 1TPO 2PTC 1PPE 1TGS 1MCT 2TBS; 2, carboxypeptidase (Ile247) **5CPA** 2CTB 1CBX 2CTC 6CPA 8CPA 7CPA 3CPA 1PCA; 3, cholesterol oxidase (Thr231) **3COX** 1COY; 4, proteinase A (Asn100) **2SGA** 3SGA 4SGA 5SGA 1SGC 3SGB; 5, citrate synthase (Glu239) **2CSC** 2CTS 4CSC 3CSC 1CSC 5CTS 1CSH 1CSI; 6, carbonic anhydrase (Lys252) **2CA2** 2CBD 2CBE 2CBA 2CBC 2CBB 3CA2 1BCD 1BIC 1CA2 1CAH 1CAI 1CAJ 1CAK 1CAM 1CAN 1CAO 1CAZ 1CIL 1CRA 1HEA 1HEB 1HEC 1HED 1MUA 1HUG 1RAY 1RAZ 1RZA 1RZB 1RZC 1RZD 1RZE; 7, subtilisin carlsberg (Asp32) **1CSE** 1SCA 1SCN 1SEL 2SEC 1S01 1S02 1ST2 1ST3 1SUB 1SUC 1SUD 2SIC 2ST1 3SIC; 8, FK506 binding protein (Ala81) **1FKF** 1FKB 1FKD 1FKG 1FKH 2FKE; 9, staphylococcal nuclease (Asn138) **1SNC** 1STN 1STY 1STA 1STB 1SNM 1ENC 1KAA 1KAB 1SYC 1SYD 1SYE 2SNM 1SYB 1SYF 2SNS; 0, glutathione reductase (His219) **3GRS** 1GRA 1GRB 1GRF 1GRE 1GRG. (b) 1, cytochrome *C*551 (Met22) 351C 451C; 2, glutathione reductase (Asn425) 3GRS 1GRA 1GRB 1GRF 1GRE 1GRG; 3, citrate synthase (Arg67) **2CSC** 2CTS 4CSC 3CSC 1CSC 5CTS 1CSH 1CSI; 4, thermolysin (Thr26) **3TLN** 6TMN 4TMN 5TMN; 5, elastase (Tyr171) **3EST** 6EST 7EST 1ELA 1ELC 1ESA 1INC 4EST 8EST 9EST; 6, retinol binding protein (Tyr111) **1RBP** 1ERB 1HBQ 1HBP; 7, D-galactose/glucose binding protein (Asp236) **2GBP** 1GCA 1GCG 1GLG; 8, ribonuclease A (Gln60) **7RSA** 1RAR 1RAS 1RAT 8RAT 1RNC 1RND 1ROB 1SRN 1SSA 1SSB 1RNU 1RNV 1RSM 2RAT 2RNS 3RAT 3RN3 4RAT 5RAT 5RSA 6RAT 6RSA 7RAT 7RSA 8RSA 9RAT 9RSA 1RBC 1RBD 1RBE 1RBF 1RBG 1RBH 1RBI; 9, proteinase K (Asp39) **2PRK** 2PKC; 0, glutathione reductase (His52) **3GRS** 1GRA 1GRB 1GRF 1GRE 1GRG. (c) 1, cytochrome *C*2 (Met91) **3C2C** 2C2C; 2, thermolysin (Ser107) **3TLN** 6TMN 5TMN 4TMN; 3, thermolysin (Asn159) **3TLN** 6TMN 5TMN 4TMN; 4, arabinose-binding protein (Asp89) **8ABP** 1APB 1ABE 7ABP 6ABP; 5, carboxypeptidase (Ser57) **5CPA** 2CTB 1CBX 2CTC 6CPA 8CPA 7CPA 3CPA 1PCA; 6, carboxypeptidase (Ser199) **5CPA** 2CTB 1CBX 2CTC 6CPA 8CPA 7CPA 3CPA 1PCA; 7, citrate synthase (His274) **2CSC** 2CTS 4CSC 3CSC 1CSC 5CTS 1CSH 1CSI; 8, thermitase (Asp38) **2TEC** 1THM 3TEC

rotation of the backbone dihedral angles (ϕ, ψ). Interestingly the ϕ angle of 5CPA Asn58 and 2CA2 Asn253 and ψ angle of 2CSC Phe66 also undergoes a large change suggesting the difference in the two classes of structures is due to a flip of the peptide. The plots of dihedral angles ϕ, ψ of 39 disallowed residues, which are conserved in related protein structures, are shown in Figure 6(a) to (d); disallowed residues are in fact ''conserved'' in structures determined by different groups of investigators.

Conclusions

The results of the present analysis demonstrate that local van der Waals contacts are overwhelmingly dominant in determining the backbone conformations of amino acid residues in globular protein structures. The number of disallowed Ramachandran non-Gly residues obtained from a carefully screened data base of 110 unrelated high resolution protein crystal structures (18,708 residues) is only 66 (0.3%). There are no examples of the occurrence of consecutive disallowed residues, a feature which may be considered in model building studies. The amino acids which most frequently adopt these disallowed conformations are Asn, Asp, Ser, Thr and His. The preponderance of small polar residues with hydrogen bonding functions in the side-chain suggests that the local compensating interactions may offset the strain energy penalty for adopting unfavorable ϕ, ψ values. Disallowed residues also form clusters in ϕ, ψ angle space suggesting that

residues adopting γ-turn conformations or those occupying the $i + 1$ position of type-II' β-turns. While no clear correlation can be drawn at present it is possible that strained backbone conformations may have a special significance at enzyme active sites. Indeed, the present data set provides several examples. The increasing number of independent protein structure determinations permits comparisons between the structures of closely related proteins and sometimes even the comparison of the same protein determined independently by different groups of investigators. Significantly, 33 disallowed residues are in fact conserved in 22 proteins where multiple structure determinations are available. While conformationally disallowed residues constitute only a very small fraction of the total number of residues in available protein structures, there are several instances where backbone stereochemical distortions are clearly an integral feature of the folded conformation. These sites may provide attractive targets for future mutagenesis studies aimed at relating local stereochemistry with energetics of folded structures.

Methods

A data set comprising 110, largely non-homologous, high resolution ($\leqslant 2$Å) protein crystal structures from the Brookhaven Protein Data Bank (PDB; Bernstein *et al.*, 1977) was examined. The data set consisted of the following PDB entries (polypeptide chain identifiers are indicated wherever homologous multiple chains are present).

1AAP A	1AK3 A	1ALC	1ALD	1BBP A	1CCR	3COX
1CRN	1CSE E	1CSE I	1CTF	1ECA	1FDX	1FKF
1FX1	1GCR	1GD1 O	1GOX	1GP1 A	1HIP	1HOE
1IFB	1MBA	1MBD	1OVA A	1PAL	1PAZ	1PCY
1PPT	1R69	1RBP	1RNH	1SN3	1SNC	1THB A
1THB B	1TON	1UBQ	1UTG	256B A	2ALP	2AZA A
2CA2	2CCY A	2CDV	2CI2 I	2CNA	2CPP	2CSC
2CYP	2ER7 E	2ER7 I	2FB4 H	2FB4 L	2FBJ H	2FBJ L
2FCR	2GBP	2HMQ A	2LH7	2LHB	2LTN A	2LTN B
2LZM	2MHR	2MLT A	2OVO	2PAB A	2PKA A	2PKA B
2PRK	2RHE	2RSP A	2SAR A	2SGA	2TEC E	5TIM A
2TRX A	2TSC A	2WRP R	351C	3APR E	3B5C	3BCL
3BLM	3C2C	3CLA	3DFR	3EBX	3EST	3FGF
3GRS	3RNT	3TLN	4BP2	4CHA A	3DFR A	4ENL
4FD1	4FXN	4I1B	4INS C	4INS D	4PEP	4PTP
4TNC	5CPA	5P21	5PTI	5RUB A	5RXN	6LYZ
6XIA	7RSA	8ABP	8DFR	9PAP	9WGA A	

specific ϕ, ψ distortions are preferred under compulsions imposed by local constraints. The clusters in the lower right-hand quadrant of the Ramachandran map are largely constituted by

Figure 1 shows the plot of ϕ, ψ values of non-Gly residues (18,708 residues) in the data set and the definition of the ''allowed region''. Region I: ($\phi = -180°$ to $-30°$, $\psi = -80°$ to $180°$); region II: ($\phi = 30°$ to $90°$, $\psi = -10°$ to $120°$);

9, erabutoxin B (Ser8) **3EBX** 5EBX 6EBX; 0, triosephosphate isomerase (Lys13) **5TIM** 1TPF 1YPI 1TPH 7TIM. (d) 1 cytochrome C2 (Thr92) **3C2C** 2*C*2C; 2, thermolysin (Thr152) **3TLN** 6TMN 5TMN 4TMN; 3, thermolysin (Ser92) **3TLN** 6TMN 5TMN 4TMN; 4, arabinose-binding protein (Asn232) **8ABP** 1APB 1ABE 7ABP 6ABP; 5, arabinose-binding protein (Pro254) **8ABP** 1APB 1ABE 7ABP 6ABP; 6, alpha-lytic protease (Ala39) **2ALP** 1PO1 1PO2 1P11 1P12 2PO7; 7, D-galactose/glucose binding protein (Asn91) **2GBP** 1GCA 1GCG 1GLG; 8, beta-lactamase (Ala69) **3BLM** 1MBL 4BLM; 9, beta-lactamase (Leu220) **3BLM** 1MBL 4BLM.

region III: ($\phi = -180°$ to $-30°$, $\psi = -180°$ to $-150°$). It should be noted that the allowed regions of Figure 1 have a slightly larger area of ϕ, ψ space (39.5%) than the regions defined using the extreme limit contact distances (25%; Ramachandran *et al.*, 1963). This slightly relaxed criterion was used for defining the stereochemically allowed regions in order to eliminate "marginally disallowed" residues and to restrict the analysis to residues having significant distortions in ϕ, ψ space. There were 146 residues having ϕ, ψ values outside these regions (0.78% of the total number of residues in the data set). The temperature factors (B-factors) of the α-carbon atom (C^α) of the chosen residues were examined and only those residues whose B-factor value is $<30Å^2$ were selected, i.e. 66 residues; 0.35% of the residues in the data set (Morris *et al.*, 1992; Tilton *et al.*, 1992). Figure 2 shows a flowchart of the scheme used in the selection of "disallowed residues".

An examination of 66 residues, selected as described above, that lie outside the broad allowed regions confirmed that disallowed residues do not occur frequently or consecutively in a protein. The 66 selected disallowed residues were distributed in 41 proteins and only nine proteins had more than two disallowed residues (seven proteins (1GD1 O, 2CSC, 2FB4 L, 3BCL, 3GRS, 5CPA, 8ABP) had three disallowed residues in each, 1GOX had four residues and 3TLN had five residues).

In early phases of this analysis, a larger data set of 113 proteins, largely non-homologous, which had been refined using crystallographic data collected to $\leqslant 2$ Å was used. However, three proteins (2SOD O, 3RP2 A, 6LDH), which had several disallowed residues even after using a B-factor filter, were removed for the following reasons: 2SOD O had 14 disallowed residues out of 151, this is significantly larger than that observed for all the other proteins in the data set. Further, several of these were either consecutive residues or proximal in sequence suggesting that the final model may have had local stereochemical problems; 3RP2 A and 6LDH had several consecutive disallowed residues, 170 to 173 in the case of the former and 207 to 208, 217 to 218 and 221 to 223 in the case of the latter. A succession of disallowed residues may indicate unacceptable stereochemistry which may be corrected by flipping connecting peptide units.

Acknowledgment

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **122**, 535–542.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392–400.

Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins: Struct. Funct. Genet.* **11**, 223–229.

Jia, Z., Vandonselaar, M., Quail, J. W. & Delbaere, L. T. J. (1993). Active-centre torsion-angle strain revealed in 1.6 Å-resolution structure of histidine-containing phosphocarrier protein. *Nature,* **361**, 94–97.

Kleywegt, G. J. & Jones, T. A. (1995). Where freedom is given, liberties are taken. *Structure,* **15**, 535–540.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

MacArthur, M. W., Laskowski, R. A. & Thornton, J. M. (1994). Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. *Curr. Opin. Struct. Biol.* **4**, 731–737.

Mattos, C., Petsko, G. A. & Karplus, M. (1994). Analysis of two residue turns in proteins. *J. Mol. Biol.* **238**, 733–747.

Milner-White, E. J. (1990). Situations of gamma-turns in proteins; their relation to alpha-helices, beta-sheets and ligand binding sites. *J. Mol. Biol.* **216**, 385–397.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Struct. Funct. Genet.* **12**, 345–364.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283–437.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.

Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.

Ramakrishnan, C. & Srinivasan, N. (1990). Glycyl residues in proteins and peptides: an analysis. *Curr. Sci.* **59**, 851–861.

Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Advan. Protein Chem.* **37**, 1–109.

Stites, W. E., Meeker, A. K. & Shortle, D. (1994). Evidence for strained interactions between side-chains and the polypeptide backbone. *J. Mol. Biol.* **235**, 27–32.

Thornton, J. M. (1992). Lessons from analyzing protein structures. *Curr. Opin. Struct. Biol.* **2**, 888–894.

Tilton, T. F., Jr, Dewan, J. C. & Petsko, G. A. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry,* **31**, 2469–2481.

Venkatachalam, C. M. (1968). Stereochemical criteria for polypeptides and proteins. Conformation of a system of three linked peptide units. *Biopolymers,* **6**, 1425–1436.

Weaver, L. H., Tronrud, D. E., Nicholson, H. & Matthews, B. W. (1990). Some uses of the Ramachandran (ϕ, ψ) diagram in the structural analysis of lysozymes. *Curr. Sci.* **59**, 833–837.

Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* **203**, 221–232.

Wilmot, C. M. & Thorton, J. M. (1990). Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng.* **3**, 479–493.

Edited by F. E. Cohen