Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Conformations of amino acids in proteins

Sven Hovmöller,* Tuping Zhou and Tomas Ohlson

Structural Chemistry, Arrhenius Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden

Correspondence e-mail: svenh@struc.su.se

Received 10 July 2001
Accepted 10 February 2002

The main-chain conformations of 237 384 amino acids in 1042 protein subunits from the PDB were analyzed with Ramachandran plots. The populated areas of the empirical Ramachandran plot differed markedly from the classical plot in all regions. All amino acids in α-helices are found within a very narrow range of φ, ψ angles. As many as 40% of all amino acids are found in this most populated region, covering only 2% of the Ramachandran plot. The β-sheet region is clearly subdivided into two distinct regions. These do not arise from the parallel and antiparallel β-strands, which have quite similar conformations. One β region is mainly from amino acids in random coil. The third and smallest populated area of the Ramachandran plot, often denoted left-handed α-helix, has a different position than that originally suggested by Ramachandran. Each of the 20 amino acids has its own very characteristic Ramachandran plot. Most of the glycines have conformations that were considered to be less favoured. These results may be useful for checking secondary-structure assignments in the PDB and for predicting protein folding.

1. Introduction

The description of conformations of amino acids in peptides and proteins by two torsion angles, φ and ψ, was proposed by Ramachandran & Sasisekharan (1968). The diagram of the two torsion angles, now known as the Ramachandran plot (Fig. 1), was first used to predict the possible conformations of the main chain. The Ramachandran plot was constructed just before the first protein structure had been determined to atomic resolution by X-ray diffraction. Based entirely on model building of short peptides, taking steric hindrance into account especially, Ramachandran predicted with a remarkable precision the conformations that could occur in proteins. As protein structures became known, it was verified experimentally that Ramachandran had correctly predicted the number of allowed areas and also the approximate numerical values of the φ, ψ torsion angles. Today, the Ramachandran plot is not only a basic diagram in textbooks on protein structures, but also a useful tool for assessing the correctness of a protein structure determination (Laskowski *et al.*, 1993; Morris *et al.*, 1992). In the first protein structures, determined to low resolution and without refinement, the φ, ψ values were rather widely scattered throughout the Ramachandran plot. However, with the advent of refined high-resolution structures, the conformation of amino acids clustered more and more near the allowed areas as predicted by Ramachandran. Today, more than 30 years after the introduction of the

Ramachandran plot, the quantity and quality of protein structures accurately determined by X-ray crystallography is such that it is possible to finally verify and slightly modify the exact shapes and positions of the allowed regions in the Ramachandran plot. That is the subject of this study.

2. Methods

We selected a non-redundant set of 1042 protein subunits from the PDB of 3 January 2002. The structures were all determined by X-ray diffraction to a resolution of 2.0 Å or higher, refined to $R \leq 0.20$ and had less than 30% sequence homology (Dunbrack, 2001). The first and last amino acids will not provide φ and ψ torsion angles. For 9940 amino acids (mainly at the beginning or end of the sequences) there were no coordinates given in the PDB. We also removed 9669 amino acids with temperature factor $B > 40$ Å^2. This left us with 237 384 amino acids for which we could calculate torsion angles.

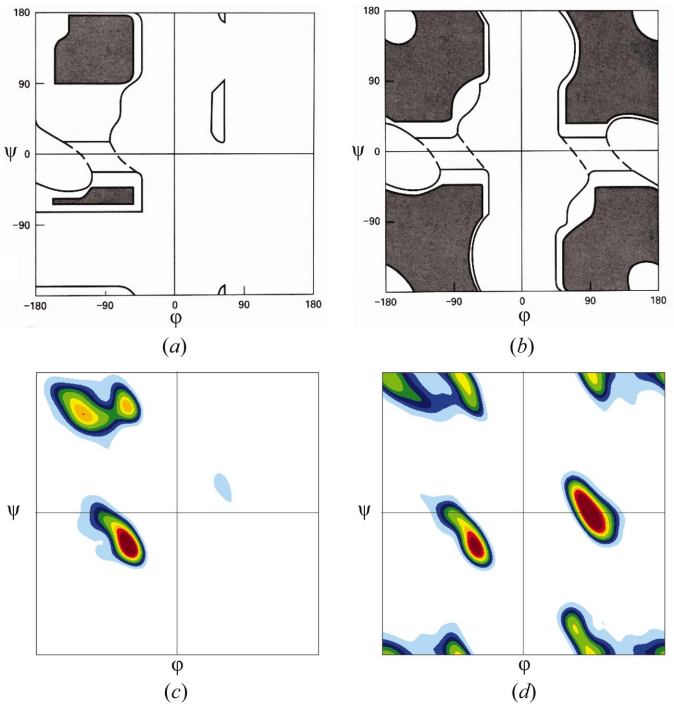


Figure 1
The classical version of the Ramachandran plot for (*a*) alanine (but often taken as typical for all non-glycines) and (*b*) glycine according to Ramachandran & Sasisekharan (1968). The fully allowed regions are shaded; the partially allowed regions are enclosed by a solid line. The connecting regions enclosed by the dashed lines are permissible with slight flexibility of bond angles. These plots were arrived at by computer modelling. Although some overall features of these plots are correct, the details differ from the experimentally observed Ramachandran plots for alanine (see Fig. 5) and (*c*) all 19 non-glycines and (*d*) glycine. The most remarkable differences are that most regions show a 45° slope rather than being parallel to any of the axes, the β-sheet region is split into two distinct maxima and the two most populated regions for glycine seen in (*d*) were predicted to be only just permissible as shown in (*b*). There are five areas in the glycine plot; two with $\psi \simeq 0°$ and three with $\psi \simeq 180°$. [(*a*) and (*b*) Reproduced from Creighton (1996) with permission.]

The amino acids were divided into three main groups according to their secondary structure as specified by *DSSP* (Kabsch & Sander, 2001). There were 88 874 in HELIX (α-helix and 3_{10}-helix) and 52 068 in SHEET (Table 1). The remaining 96 442 amino acids for which atomic coordinates were given in the PDB but were not specified as HELIX or SHEET by *DSSP* are called Random coil in this investigation, although this group of course includes turns and other well defined structures. Ramachandran plots for each of the 20 amino acids were made for HELIX, SHEET, Random coil and All, where All includes all the different secondary structures.

The Ramachandran plot was split into four regions for detailed analysis of the conformations. Amino acids with torsion angles in the range $-180 < \varphi < 0°$, $-100 < \psi < 45°$ were considered to be in the α-helical region. Amino acids with torsion angles in the range $-180 < \varphi < -45°$, $45 < \psi < 225°$ were considered to be in the β-sheet region. The area $0 < \varphi < 180°$, $-90 < \psi < 90°$ is here called the turn region. The remaining area of the Ramachandran plot represents 36% of the area but contained only 1.9% of the amino acids and was not studied further. The bridging region between α-helical and β-sheet conformation is often given as $-135 < \varphi < -45°$, $-25 < \psi < 15°$, but we found that this region should really be considered α-helical and the bordering region should be moved upwards to $-160 < \varphi < -65°$, $45 < \psi < 90°$.

Amino acids in SHEET were subdivided into six groups according to their contacts with adjacent strands (Fig. 2). These were (i) parallel with one partner, (ii) antiparallel with one partner, (iii) parallel with two partners, (iv) antiparallel with two partners, (v) parallel with one and antiparallel with another partner and (vi) amino acids without partner.

The amino acids in HELIX were subdivided in several ways. There were 1907 three-residue-long helices and 1023 four-residue-long helices and these were treated separately. For the other α-helices, we looked specifically at the first and last amino acids. In order to find the ideal α-helical conformation, all but the first two and last two amino acids in helices ≥ 5 amino acids long were pooled.

Amino acids in Random coil were first subdivided into the four areas of the Ramachandran plot: α-helical, β-sheets, turns and others. Within each of these areas, the stretches of random

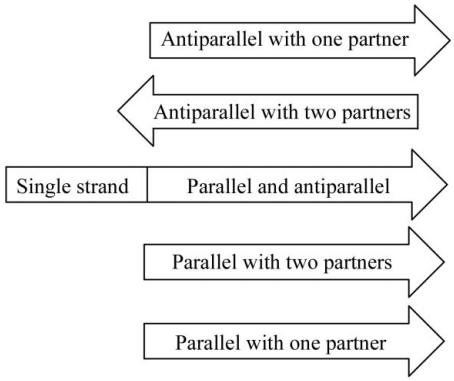


Figure 2
The definitions of six types of β-strands.

coil were sorted according to their lengths. Long stretches of amino acids in Random coil all in α-helical or all in β-sheet conformation were checked one by one. The longest stretches of amino acids all in our generously defined α-helical region but not defined as α-helical by *DSSP* were five or six. When we made a similar search using the secondary-structure assignment in the PDB, we found many α-helices longer than 20 residues. There were many single strands, *i.e.* long stretches of amino acids all in the β-sheet area but not assigned as SHEET by *DSSP* or PDB.

The Ramachandran plots were produced in the following way. Firstly, all amino acids used were sorted into 90 × 90 squares, each covering an area of 4 × 4°. The number of amino acids in each such square was summed. The squares were sorted from the highest abundances to the lowest. Each square was colour-coded such that each colour represented 10% of all amino acids (Fig. 3). In order to make the plots easier to interpret, they were smoothed. The effect of smoothening can be appreciated by comparing the plot before and after smoothing, as shown for one of the most complex cases, Asp in Random coil (Fig. 4).

3. Results

As expected, most of the amino acids (99%) specified as HELIX by *DSSP* were found in the α-helical region of the Ramachandran plot. Similarly, 95% of all amino acids specified as SHEET by *DSSP* were found in the β-sheet region. Those amino acids that were found in unexpected areas were often the first or last amino acid in a strand or in an α-helix; many were glycines.

The Ramachandran plots of glycine and the other 19 non-glycine amino acids show several interesting deviations from the classical Ramachandran plot (see Fig. 1). There are two distinct maxima in the β-sheet region, as has already been pointed out by Kleywegt & Jones (1996). Here, we will analyze the features of the different regions in the Ramachandran plot in greater detail.

Each of the 20 amino acids has its characteristic Ramachandran plot (Fig. 5). Perhaps the most remarkable of these is Gly, which differs very much from the classical Ramachandran plot: the two dominating areas are in regions thought to be only just possible. There are also many glycines around $\varphi = \psi = \pm 180°$, an area thought to be disallowed. All amino acids except Gly and Pro have the highest density of points in the α-helical area. Most amino acids have two distinct distributions in the β-sheet region. The only exceptions are Ile, Pro and Val. Val and Ile have nearly identical distributions and are also similar in their unique property of preferring β-sheets to α-helix. The three large hydrophilic amino acids Phe, Tyr and Trp have quite similar distributions. Phe and especially Tyr also show two distinct maxima in the α-helical region. Asn and Asp have quite similar Ramachandran plots, but differ from all other amino acids in having a very complicated β region and the next highest frequency of amino acids in the turn region after Gly. They are also more frequently found in the bridging region between α-helix and β-sheet than any of the other amino acids.

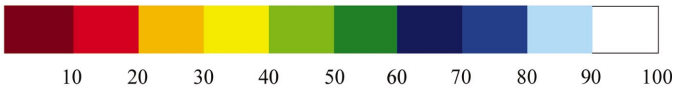


Figure 3
The colour table used throughout this paper. The area in each colour contains 10% of all the amino acids in the respective plot. The dark red colour marks the most densely occupied regions of the Ramachandran plot. The most sparsely populated region in white typically covers 80–90% of the total area, but contains only 10% of all amino acids.

Table 1
The amino acids sorted into secondary structure as specified by *DSSP* and according to their torsion angles.

	HELIX (*DSSP*)	SHEET (*DSSP*)	Random coil (*DSSP*)	Sum
α-Helical	88155	1556	30745	120456
β-Sheets	143	49370	51813	101326
Turns	283	267	10537	11087
Others	293	875	3347	4515
Sum	88874	52068	96442	237384

When the amino acids were split into the three groups of secondary structure, HELIX, SHEET and Random coil, more details appeared (Figs. 6, 7, 8 and 9). As expected, only very few amino acids considered to be parts of β-sheets had conformations in the α-helical region and *vice versa* (see Table 1). Most of those amino acids were found at the ends of strands or helices and may be considered either as highly distorted or sometimes simply incorrectly assigned by *DSSP*.

3.1. The effects of resolution and temperature factors

When the subunits were sorted into different resolution ranges, the results were similar. The only effect of increased resolution was that the distributions of φ/ψ angles sharpened, just as observed by Walther & Cohen (1999). We also analysed the effect of deleting all amino acids with high temperature factors ($B > 40$ Å^2). As expected, the hydrophilic amino acids, which are more often found at the protein surfaces, are somewhat over-represented among those with $B > 40$ Å^2. The

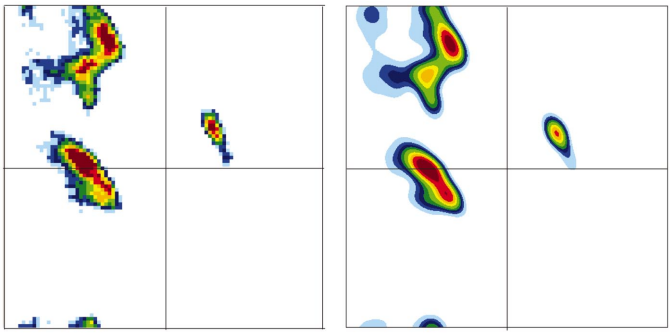


Figure 4
The effect of smoothing is demonstrated here for one of the most intricate Ramachandran plots: Asp in Random coil. (*a*) Before smoothing, (*b*) after smoothing.

average temperature factor for all amino acids was 18.4 Å^2 (17.0 Å^2 for those with $B \leq 40$ Å^2). About 4% of all amino acids had B values > 40 Å^2. The average B values were calculated as the average B of the four atoms (C, C$^\alpha$, O and N) which are used for calculating the φ/ψ torsion angles.

3.2. The α-helical region

For every one of the 20 amino acids the experimentally observed φ, ψ torsion angles in the α-helical region are distributed along a line parallel to $\varphi = -\psi$ (Figs. 1*c* and 5), as already pointed out by Garnier & Robson (1990). This does not agree with the areas defined as fully and partially allowed by Ramachandran & Sasisekharan (1968). They predicted that there should be a large range of free rotation about the φ torsion angle (-160 to -50°), while the rotation around ψ was expected to be very limited (-60 to -40°). It is clear from the experimentally determined protein structures that the allowed rotations about φ and ψ are equally large and that these rotations are coupled. Thus, if φ is rotated by, for example, $+30^\circ$, then ψ is also rotated by -30°. This coupling of the two torsion angles φ and ψ arises from interactions between the carbonyl O atom in amino acid n with both the carbonyl O atom in amino acid $n + 1$ and the backbone N atom in amino acid $n + 2$. These O—O and O—N distances are always kept close to 3.5 Å.

When five or more consecutive amino acids have torsion angles in the α-helical region of the Ramachandran plot, the hydrogen bonds typical of the α-helix will form and the structure will snap into a very highly populated and finely focused area near $\varphi = -63.8$, $\psi = -41.1^\circ$ (see Fig. 6*e*). As many as 39% of all 237 384 amino acids were found here, in an area just 2% of the Ramachandran plot ($-89 < \varphi < -39^\circ$, $-66 < \psi < -16^\circ$). The data points that form the elongated distribution parallel to $\varphi = -\psi$ come from amino acids in Random coil (Fig. 9), from three-residue- and four-residue-long helices and from the last amino acid in α-helices (Figs. 6*a*, 6*b* and 6*d*).

The conformations of each of the 20 amino acids inside α-helices of length at least five (*i.e.* skipping the first and last two amino acids in each α-helix) were investigated. All 20 amino acids have torsion angles within $\pm 2^\circ$ of the average value $\varphi = -63.8$, $\psi = -41.1^\circ$, except Pro and Gly which are outliers by a few degrees (Pro $-61.0/-36.5^\circ$, Gly $-59.1/-42.4^\circ$). The differences in average torsion angles between

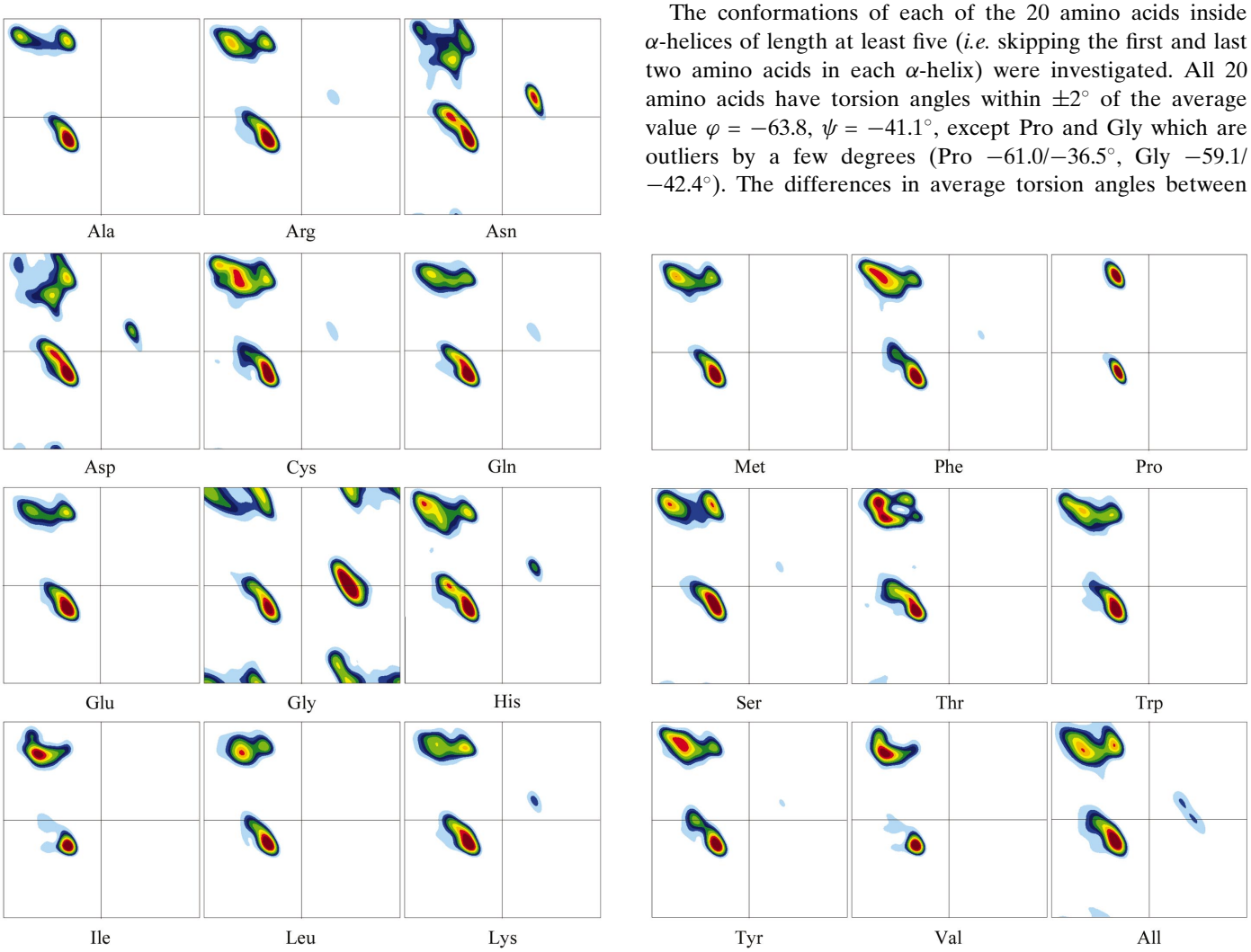


Figure 5
Individual Ramachandran plots for each of the 20 amino acids (All includes all 20 amino acids). Notice that most amino acids have two distinct maxima in the β-sheet region (upper left quadrant). Asp and Asn have the most complicated plots after Gly. This reflects their role in terminating α-helices and β-sheets. The two amino acids with highest preference for β-sheets, Ile and Val, have very similar Ramachandran plots. The plots of the three large hydrophobic amino acids Phe, Tyr and Trp look alike.

amino acids in α-helices are much less than one standard deviation of any individual amino acid (5–15°).

3.2.1. Amino acids in Random coil with α-helical conformations. Nearly one third of all amino acids in Random coil have torsion angles in the α-helical region (Table 1). However, the distribution of these torsion angles is strikingly different from that of amino acids in α-helices. While the torsion angles inside α-helices are highly focused near the point $\varphi = -63.8$, $\psi = -41.1°$, the amino acids in Random coil have torsion angles along a line stretching from $\varphi = -50$, $\psi = -40°$ to $\varphi = -120$, $\psi = +30°$ (Fig. 9).

3.3. The β-sheet region

The upper left corner of the Ramachandran plot contains amino acids in the β-sheet conformation (Table 1). The distribution of points continues above $\psi = +180°$, so we included all data in the ranges $-180 < \varphi < -45°$, $+45 < \psi < +225°$ in our investigations of the β-sheet conformations. When all 19 non-glycine amino acids are pooled (Fig. 1*c*), the β area is split into two distinct regions. The widest and most populated region is elongated with a centre at $\varphi = -121$, $\psi = +128°$, while the other region is nearly circular around $\varphi = -66$, $\psi = +137°$. However, each individual amino acid has its special variation on this common theme, as can be seen in Figs. 5, 8 and 9.

3.3.1. Single strands. Nearly 43% of all amino acids have β conformation. More than half of these (51%) are not parts of strands in β-sheets but occur in Random coil. In comparison with the α-helices, the situation is more complex for β-sheets. If five or more consecutive amino acids have α-helical conformation, they form an α-helix. In contrast, some long stretches of ten or more consecutive amino acids with β conformation are not called SHEET in the PDB or by *DSSP* because they do not have at least one parallel or antiparallel strand as a partner to which they are connected through hydrogen bonds. It is a commonly accepted dogma that 'the β-strand conformation is stable only when incorporated into a β-sheet'. However, we have found hundreds of single strands with four or more consecutive amino acids all with β conformation, yet not denoted SHEET either in the PDB file or by *DSSP*. Many of these are indeed single β-strands. The longest single β-strand we found was in 1bfd, amino acids 334–350.

3.3.2. Parallel and antiparallel β-strands. The two types of β-sheet, parallel and antiparallel, have quite different hydrogen-bond connections. It has been generally believed that the conformations of the amino acids in these two types of β-sheet should be quite different. In order to analyze the distributions of torsion angles for amino acids in β-sheets, we sorted them into six categories. As is evident from Fig. 2, it is not sufficient to use only the two groups parallel and antiparallel; many β-strands have one parallel and one antiparallel strand and those β-strands that are at the outside of a β-sheet, *i.e.* have only one partner, do not necessarily have the same conformations as those inside a β-sheet. Finally, there were quite a number of amino acids denoted SHEET in the PDB and/or *DSSP* even though they were not specified as being connected by a hydrogen bond to another strand. These were usually (but not always) the first or last amino acids in correctly denoted β-strands and constitute the sixth and last group in our studies of β-sheets.

The conformations of the amino acids in five of the six groups were surprisingly similar (Fig. 7) with respect to both the average φ and ψ values (all within $-130 < \varphi < -105°$,

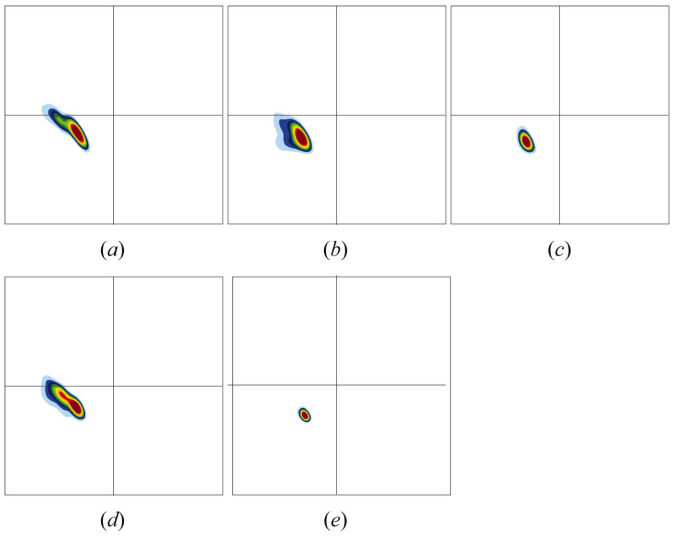


Figure 6
Ramachandran plots for amino acids marked as HELIX in *DSSP*: (*a*) three-residue-long α-helices, (*b*) four-residue-long α-helices, (*c*) the first amino acid in α-helices of any length, (*d*) the last amino acid in α-helices of any length and (*e*) inside at least five-residue-long α-helices, with the first two and last two amino acids in each α-helix removed.

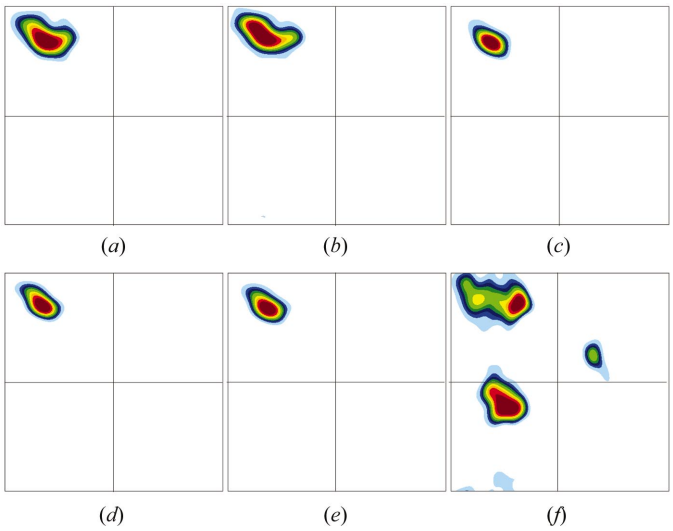


Figure 7
Ramachandran plots for amino acids marked SHEET in *DSSP*, grouped into the six types of β-strands as defined in Fig. 2: (*a*) parallel with one partner, (*b*) antiparallel with one partner, (*c*) parallel with two partners, (*d*) antiparallel with two partners, (*e*) parallel with one partner and antiparallel with another partner and (*f*) amino acids without partners. Notice that the first five groups are quite similar, illustrating that the conformations of parallel and antiparallel β-sheets are nearly the same. Clearly, the amino acids in β-strands contribute mainly to the leftmost of the two maxima in the β-sheet region seen in Fig. 1(*c*).

$128 < \psi < 147^\circ$) for the first five groups and the standard deviations (15–26°). Most of the amino acids were found in the upper left corner of the β region. Especially remarkable was that the differences between the parallel and antiparallel β-strands were much less than expected (Figs. 7*c* and 7*d*). The average torsion angles for the 18 amino acids excluding Pro and Gly were $\varphi = -122$, $\psi = +136^\circ$ for antiparallel and $\varphi = -116$, $\psi = +128^\circ$ for parallel β-sheets. Typical values for torsion angles given in the literature are $\varphi = -139$, $\psi = +135^\circ$ for antiparallel and $\varphi = -119$, $\psi = +113^\circ$ for parallel β-sheets. Thus, the φ angles differ only by about 6° compared with the expected 20° and the ψ angles differ by 8° rather than the expected 22°. Those internal strands that have one parallel and one antiparallel partner (Fig. 7*e*) are very similar to the β-strands with two similar partners (Figs. 7*c* and 7*d*). Strands with only one partner (Figs. 7*a* and 7*b*) are similar to strands with two similar partners (Figs. 7*c* and 7*d*) but are more spread along φ and ψ. The standard deviations are $\varphi = 25$, $\psi = 20^\circ$ for strands with one partner and $\varphi = 17$, $\psi = 15^\circ$ for strands with two partners.

Finally, those strands that do not have a partner although they are denoted SHEET by *DSSP* are distributed over all three main areas of the Ramachandran plot (Fig. 7*f*). For all groups, the upper right corner of the β region ($\varphi > -90$, $\psi > 135^\circ$) has a very high proportion of prolines: 20% for all six groups taken together.

The amino acids in the strands were also split into three groups depending on where they occurred in the strands; first, last or inside. These groups did not differ very much. On average, the first amino acids in SHEET have 4° larger ψ angles than those inside the β-strands and at the ends. In most cases, the torsion angles differ only by a degree or two for averages calculated using all amino acids in the strands or excluding the first and last amino acid in each strand.

Furthermore, when we looked at each of the 20 amino acids separately, the trends mentioned above were the same. All amino acids except Pro have on average more negative φ angles and more positive ψ angles for antiparallel than for parallel β-strands (Table 2). The φ and ψ angles of Thr are equal to within a degree in parallel and antiparallel strands. As expected, Pro and Gly are outliers. All the other 18 amino acids have average ψ angles for antiparallel ranging from +129° (Asp) to +145° (Ser), with an average of $\psi = +136^\circ$. The φ values range from −113° (Asp) to −130° (Ala), with an average of $\varphi = -122^\circ$. It is clear that the ψ value +135° found

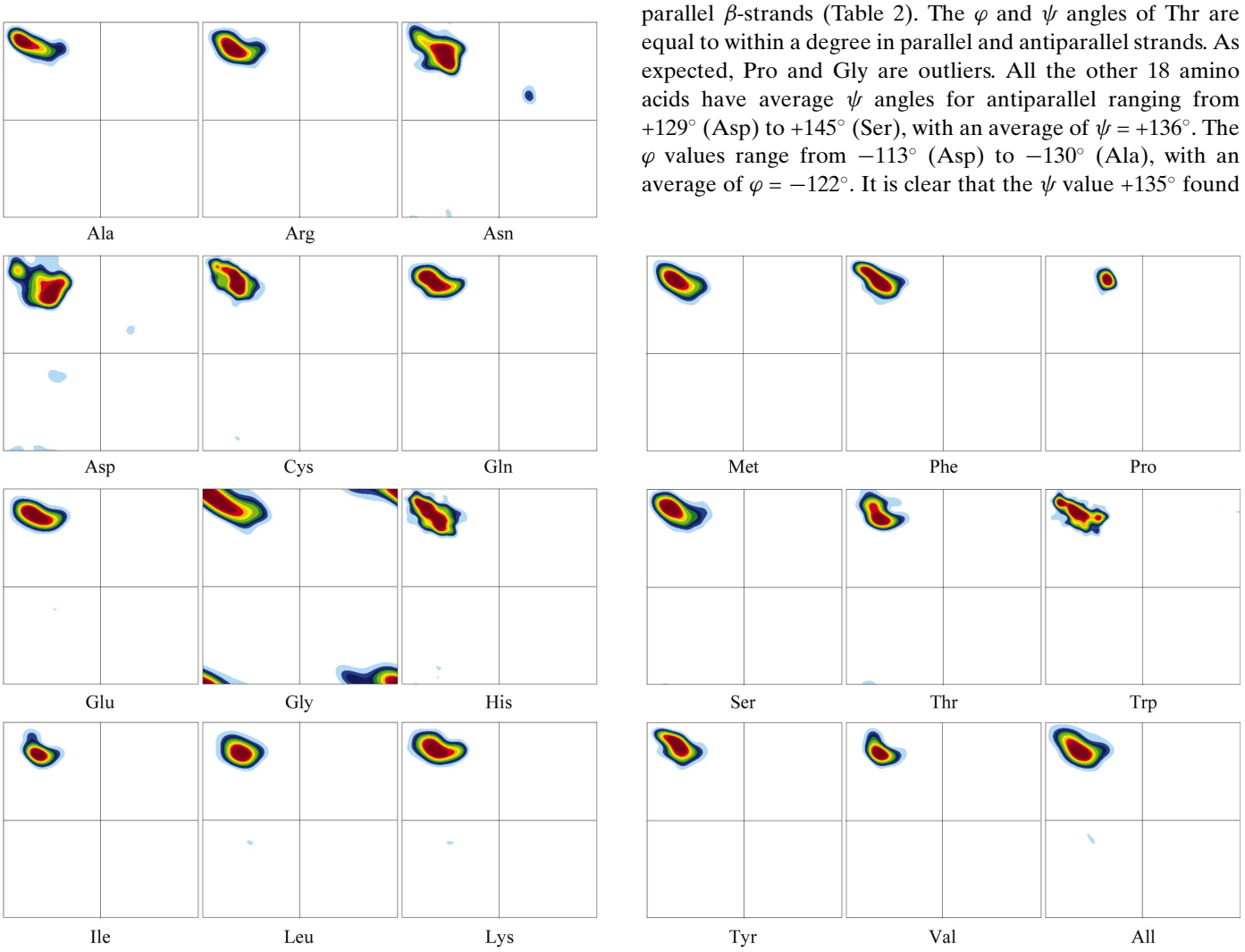


Figure 8
Individual Ramachandran plots for each of the 20 amino acids in SHEET according to *DSSP* (All includes all 20 amino acids). Glycine has its maximum centred at $\varphi = 180$, $\psi = 180^\circ$, one of the few conformations predicted to be disallowed in the original Ramachandran plot, as seen in Fig. 1(*b*). Proline has a conformation quite distinct from the remaining 18 amino acids, which otherwise only display minor differences between themselves.

in the literature is correct, while the theoretical φ angle -139° is definitely too negative.

For parallel β-strands, the average torsion angles range from $\varphi = -111^\circ$ (Asp) to -123° (Thr) with an average of $\varphi = -116^\circ$. The ψ values range from $+119^\circ$ (Asp) to $+137^\circ$ (Thr) with a total average of $\psi = +128^\circ$. In this case the theoretical φ angle (-119°) is acceptable, but the ψ angle ($+113^\circ$) is far too small. The theoretical differences between antiparallel and parallel φ/ψ angles are 20 and 22°, respectively. The largest difference we found for an individual amino acid was 13° in φ for Trp and 15° in ψ, also for Trp. For half of the amino acids, the average φ and ψ angles differed by 7° or less between parallel and antiparallel strands (Table 2). It is evident that the conformations of amino acids in parallel and antiparallel strands do not differ as much as hitherto believed. The spread of torsion angles for β-sheets is much larger than for α-helices. The estimated standard deviations of φ and ψ angles are between 12 and 22° for each amino acid (Table 2), which is much more than the average differences between parallel and antiparallel β-sheets. Thus, the populations of parallel and antiparallel β-sheets overlap to a very large extent.

Table 2
The average φ, ψ values (aver) and the estimated standard deviations (e.s.d.) for parallel and antiparallel β-strands and the differences (parallel − antiparallel).

	Parallel with two partners					Antiparallel with two partners					Differences	
	φ_{aver}	$\varphi_{e.s.d.}$	ψ_{aver}	$\psi_{e.s.d.}$	No.	φ_{aver}	$\varphi_{e.s.d.}$	ψ_{aver}	$\psi_{e.s.d.}$	No.	$\Delta\varphi$	$\Delta\psi$
Ala	−122.0	22.0	136.6	16.8	376	−130.2	21.4	143.8	14.6	753	8.2	−7.1
Val	−117.7	12.9	127.8	11.6	1209	−121.2	13.6	132.5	13.2	1 470	3.6	−4.7
Leu	−111.6	15.8	124.9	12.9	806	−115.2	15.8	131.8	13.5	1141	3.6	−7.0
Ile	−115.4	12.9	125.7	11.7	953	−118.5	13.5	130.5	12.4	1034	3.2	−4.8
Phe	−113.9	18.5	129.3	17.1	333	−123.3	19.4	140.6	16.9	640	9.4	−11.4
Met	−116.2	18.4	129.0	16.3	149	−127.5	18.3	139.3	14.4	269	11.3	−10.3
Lys	−119.1	17.3	130.9	13.1	67	−118.9	17.5	134.2	15.3	382	−0.3	−3.3
Arg	−116.4	19.5	128.7	15.4	81	−123.9	17.1	135.9	14.3	389	7.5	−7.2
His	−113.9	20.3	124.4	17.7	89	−122.4	20.6	136.0	18.6	187	8.5	−11.6
Ser	−119.2	20.1	134.6	16.0	153	−129.7	19.0	145.1	15.2	453	10.6	−10.5
Thr	−122.6	15.3	137.0	16.6	229	−123.9	14.0	138.1	14.3	602	1.3	−1.1
Cys	−122.0	17.0	131.2	19.1	79	−124.2	17.4	139.3	18.1	202	2.2	−8.2
Tyr	−119.3	16.9	131.5	16.0	220	−125.5	18.4	142.8	15.6	710	6.3	−11.3
Asn	−112.3	18.7	123.8	17.2	87	−117.9	20.3	135.7	17.4	196	5.6	−11.9
Glu	−116.7	19.3	128.9	14.1	78	−121.0	18.2	134.4	13.1	384	4.3	−5.5
Trp	−112.1	16.7	127.4	15.6	71	−124.7	19.8	142.7	15.2	235	12.7	−15.3
Asp	−111.3	19.1	118.7	17.0	91	−113.4	19.7	129.0	19.9	195	2.1	−10.4
Gln	−114.5	16.2	129.4	15.5	77	−124.5	19.0	136.9	15.2	265	10.1	−7.4
Average	−116.3	16.4	128.4	14.6	5148	−122.2	17.7	136.4	15.5	9507	5.9	−8.0

3.3.3. Ramachandran plots of individual amino acids in β-sheets. The individual Ramachandran plots for each of the 20 amino acids designated SHEET according to *DSSP* are shown in Fig. 8. Glycine is very different from the others; it has one broad maximum centred at $\varphi = 180$, $\psi = 180^\circ$, one of the few conformations predicted to be disallowed for glycine in the original Ramachandran plot, as seen in Fig. 1(*b*). Proline has a conformation quite distinct from the remaining 18 amino acids; it is highly focused and has a less negative φ angle than any of the others. The remaining 18 amino acids are all rather similar. They all have a single broad maximum, elongated mainly in the φ angle. This maximum corresponds to the leftmost of the two maxima found in the β-sheet region (see Fig. 1*c*).

3.3.4. Amino acids in Random coil with β-sheet conformations. Only 49% of all amino acids with β-conformation are specified as SHEET by *DSSP* (Table 1). A few (0.1%) are specified as HELIX, while most (51%) are found in segments of Random coil. The latter were analyzed with respect to φ and ψ angles for each of the 20 amino acids and as a function of length of the single strands.

The Ramachandran plots for all 20 amino acids in Random coil (Fig. 9) with a conformation in the β-sheet region have several significant differences compared with those of the amino acids in SHEET (Fig. 8). Most of the amino acids in Random coil (55%) but only 22% of those in SHEET are found in the upper right corner of the β region. The average φ angle in Random coil is about 22° larger than the average φ-angle value encountered in β-sheets, while the ψ angle is not much different from that in β-sheets. There were no significant differences in torsion angles as a function of strand length. The torsion angles are spread over a wide range of φ angles, from -180 to -40°, but the highest density of Random coil points is in the upper right corner of the β region whereas the highest density of SHEET points in the Ramachandran plot is in the left region. One reason for the difference between the conformations in β-sheets and Random coil is that there are many prolines (8%) in Random coil but very few (2%) in β-sheets. All prolines, with their rather rigid conformation, are found in the upper right corner of the β region. However, the prolines do not account for more than 20% of the amino acids in Random coil in that region, so there must be other reasons for the differences in conformations in Random coil and SHEET. For every one of the 20 amino acids, the average φ angle is less negative when they occur in Random coil than when they occur in SHEET, typically by 15–25°.

The ψ angles do not differ very much between amino acids in Random coil and in SHEET (3°). However, in Random coil there are two bridges between the β-sheet region and the α-helical region. One bridge is around $\varphi = -130^\circ$ and the other around $\varphi = -80^\circ$. Both bridges have ψ values ranging from $+45$ to $+90^\circ$. Asp and Asn are most common in these areas; they make up about 27% of the amino acids in these two bridges. The two bridging regions depend in a remarkable way on proline; nearly 60% of the amino acids in the left bridging region are followed by Pro, while none in the right bridging region is followed by Pro.

The three areas in each corner of the Ramachandran plot that can be considered as a continuation of the β-sheet region are all mainly populated by Gly (see Fig. 5). This is especially true for the area $135 < \varphi < 180^\circ$ and $-180 < \psi < -135^\circ$; 576 out of the 579 amino acids found in this region were Gly!

3.4. The turns region

The last and least-populated region of the Ramachandran plot, the turn region (also called the left-handed α-helical region) around $\varphi = \psi = +60^\circ$, is mainly populated by glycines

in turns. Only 4% of all amino acids have torsion angles in the range $0 < \varphi < 180^\circ$, $-90 < \psi < 90^\circ$. Of these 58% are Gly, 11% Asn and 6% Asp. There are almost no Pro and only very few Val and Ile in this region.

The distribution of points in the Ramachandran plot in the turn region is elongated parallel to the line $\varphi = -\psi$, *i.e.* parallel to the distribution of points in α-helical conformation. This distribution again differs from that suggested in the original Ramachandran plot, where the φ angle was confined to a narrow range around 50° while the ψ angle could vary from +20 to +100°. However, it is clear from the present investigation that the two torsion angles φ and ψ are coupled here, just as they are for amino acids in the α-helical region. Furthermore, the turn region is situated below the predicted left-handed α-helical region.

A very interesting picture emerges when the glycines are plotted separately (Fig. 10*a*) and compared with all the other amino acids in this region (Fig. 10*b*). These two distributions are almost non-overlapping, with the glycines mainly having $\psi < 25^\circ$ and the other amino acids having $\psi > 25^\circ$. The non-glycines are quite well focused near $\varphi = 60$, $\psi = 35^\circ$, a point mirroring the α-helical conformation $\varphi = -64$, $\psi = -41^\circ$. Apart from glycine, all the other 19 amino acids have quite similar distributions of torsion angles in this region.

It has been proposed that there could exist a left-handed α-helix with torsion angles near $\varphi = 60$, $\psi = 60^\circ$. We only found 71 stretches of three, three of four and one stretch of five consecutive amino acids in this region. About half of these are just indicated as HELIX in the PDB files, while in a few cases the authors have explicitly marked the helix as left-handed. In some cases, the left-handed helix is directly connected to a normal right-handed α-helix, for example 1afw (residues 97–99), but in the PDB file the whole sequence 96–106 is marked only as HELIX. Our results are in total agreement with those of Kleywegt (1999), who also found only one protein with a five-residue-long left-handed α-helix (in 1sft which is also included in our sample). Only 1% of the 11 087 amino acids found in this region of the Ramachandran plot can be considered as being parts of left-handed α-helices with three or more amino acids. For this reason, we discourage the use of this name for this region of the Ramachandran plot. We propose to call it the turn region, since most of the amino acids

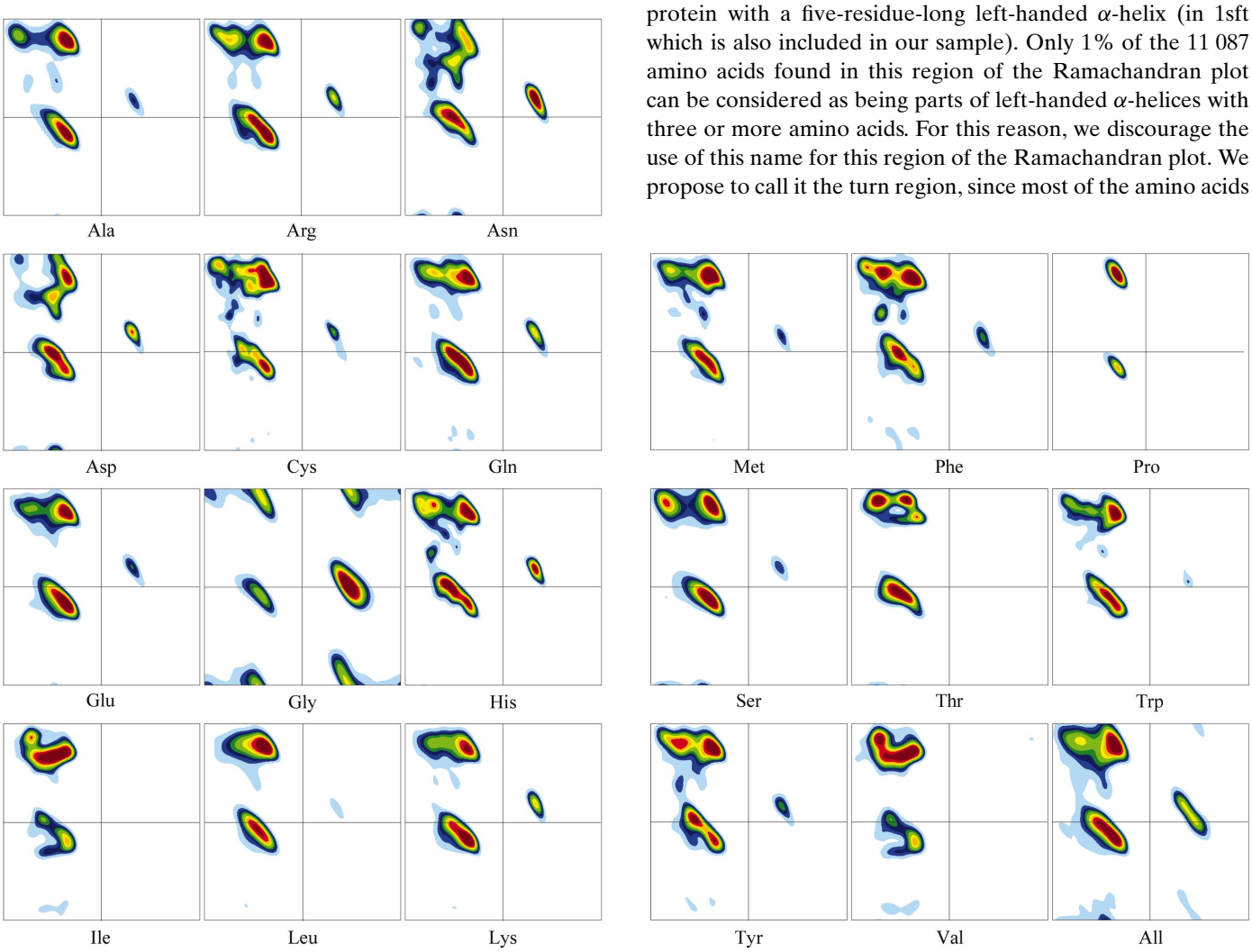
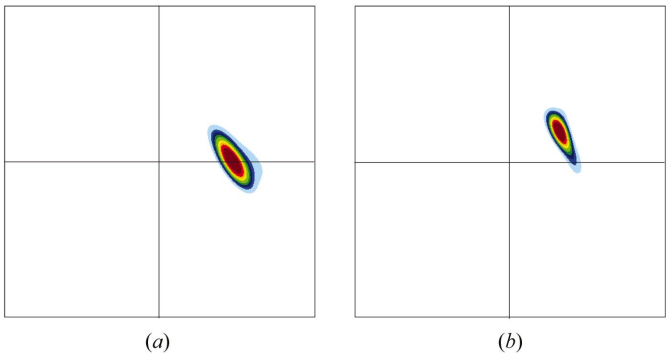


Figure 9
Individual Ramachandran plots for each of the 20 amino acids in Random coil, *i.e.* neither denoted HELIX nor SHEET in *DSSP* (All includes all 20 amino acids). All amino acids except Pro, Ile, Val and Thr have a significant proportion of amino acids in the turn region. The amino acids in Random coil contribute mainly to the rightmost of the two maxima in the β-sheet region seen in Fig. 1(*c*). The amino acids in Random coil in the α-helical part of the Ramachandran plot are mainly located above and to the left of the conformation inside α-helices as seen in Fig. 6(*e*). Many amino acids have conformations in two bridging regions between β-sheets and α-helices, as is especially clear in the case of Asn.



(*a*) (*b*)

Figure 10
Ramachandran plots for amino acids in the turn region. More than half the amino acids in the turn region are Gly. Their distribution is shown in (*a*). The conformations of all the other amino acids are quite similar and are merged in (*b*). The distribution of Gly is below and almost non-overlapping with that of the other amino acids.

found here are parts of turns. In retrospect, we think it would have been much better to call the three most populated areas of the Ramachandran plot the α-helix, β-strand and γ-turn regions.

4. Discussion

We have found several interesting and perhaps surprising features in protein structures as the result of this extensive analysis of Ramachandran plots. The β-sheet region is especially interesting, with its complicated fine structure differing from one amino acid to another. Most amino acids have two distinct narrow regions of φ, ψ angles. Some, especially Asn and Asp, have even more complex Ramachandran plots. Although many of these features can be discerned already in the plots by Kleywegt & Jones (1996), they are more evident here owing to our much larger high-resolution sample and the smoothing procedure.

One surprise was the frequent observation of isolated single β-strands, sometimes over ten amino acids long. When the β-sheet was proposed as a possible secondary structure for proteins, it was observed that two or more parallel or antiparallel β-strands could be joined by a set of hydrogen bonds, which of course would stabilize the structure. However, this does not rule out the possibility that single β-strands can also be stabilized by hydrogen bonds in other less regular ways than as parts of β-sheets.

There is only a very small difference in conformations of amino acids that are parts of parallel or antiparallel β-sheets. The differences between these two groups are only 6 and 8° in each of the φ and ψ angles, respectively, compared with the 20 and 22° in each angle expected from model building. Considering the rather high standard deviations of the torsion angles, 13–22°, the difference in torsion angles between parallel and antiparallel strands are almost insignificant. However, it should also be mentioned that for nearly all amino acids the average φ values are more negative and the ψ values larger for the antiparallel than for the parallel strands. Thus, the trend is the same as that expected, but the differences between parallel and antiparallel are much smaller than expected.

The Ramachandran plot is a very useful tool for checking protein structure determinations. Much more detailed checks than is the practice today may reveal many more errors. Any stretch of four or more consecutive amino acids in α-helical conformations but not marked HELIX should be checked carefully and will most probably be found to be an α-helix. Similarly, stretches of four or more consecutive amino acids with β conformation but not assigned as SHEET should be checked. Any single amino acid in HELIX outside the α-helical region or SHEET outside the β-sheet region must also be checked.

Our findings may also prove important for protein-folding prediction. For that purpose, it is imperative to consider the relevant clusters of conformations. The different populations in the β-sheet region must be treated and analyzed separately.

A list of the 1042 protein subunits used in this study can be obtained from the authors.

We thank Salam Al-Karadaghi and Michel N. Fodje for valuable discussion.

References

Creighton, T. E. (1996). *Proteins: Structures and Molecular Properties*, 2nd ed., pp. 172–176, 183. New York: W. H. Freeman & Company.
Dunbrack, R. (2001). *Culling the PDB by Resolution and Sequence Identity.* http://www.fccc.edu/research/labs/dunbrack/culledpdb.html/
Garnier, J. & Robson, B. (1990). *Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. Fasman, pp. 417–465. New York & London: Plenum Press.
Kabsch, W. & Sander, C. (2001). *The DSSP Database.* ftp://www.cmbi.kun.nl/pub/molbio/data/dssp/.
Kleywegt, G. J. (1999). *J. Mol. Biol.* **285**, 1887–1897.
Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.
Walther, D. & Cohen, F. E. (1999). *Acta Cryst.* D**55**, 506–517.