

Oliviero Carugo<sup>a,b\*</sup> and Kristina  
Djinović-Carugo<sup>b,c\*</sup>
<sup>a</sup>Department of Chemistry, University of Pavia,  
Viale Taramelli 12, I-27100 Pavia, Italy,

<sup>b</sup>Department of Structural and Computational  
Biology, Max F. Perutz Laboratories, University  
of Vienna, Campus Vienna Biocenter 5,  
A-1030 Vienna, Austria, and <sup>c</sup>Department of  
Biochemistry, Faculty of Chemistry and  
Chemical Technology, University of Ljubljana,  
Aškerčeva 5, SI-1000 Ljubljana, Slovenia

 Correspondence e-mail:  
oliviero.carugo@univie.ac.at,  
kristina.djinovic@univie.ac.at

## Half a century of Ramachandran plots

On the occasion of their fiftieth birthday, it is opportune to review the first half century of Ramachandran plots. In the present review, some of the most relevant aspects of this fifty-year history are summarized, from the original ideas of Gopalasamudram Narayana Ramachandran to subsequent revisions and to applications in structural biology. This will not be a guided walk through five decades of Ramachandran plots, but a commented summary of the lines along which the original ideas evolved and continue to develop, and of their applications.

Received 28 February 2013

Accepted 27 April 2013

## 1. Original idea

In 1963, Gopalasamudram Narayana Ramachandran and his coworkers were able to predict which conformations of the polypeptide backbone are possible by using simple electric desk calculators (Sasisekharan, 1962; Ramachandran *et al.*, 1963; Ramachandran & Sasisekharan, 1968; Ramakrishnan & Ramachandran, 1965). In retrospect, these predictions were fully in line with the experimental protein structure determination of myoglobin in 1958 (Kendrew *et al.*, 1958).

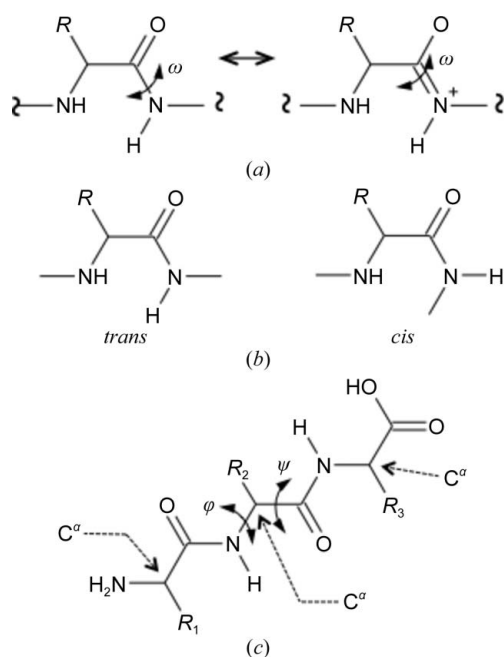
A few clever assumptions made the original predictions possible. There are four covalent bonds in the protein backbone. One of them, the carbonylic C=O double bond, is irrelevant from a stereochemical perspective. Rotations around it are impossible and even if they did occur they would not affect the shape of the polypeptide backbone. The relevance of the bond between the carbonylic C atom and the amidic N atom of the next amino acid is also minor, since there are only two possible geometries. Given the partial double-bond character of this carbon–nitrogen bond, this bond can be in a *cis* or a *trans* conformation, which means that the dihedral angle  $\omega$  can only assume a value of 0° (*cis*) or a value of 180° (*trans*) (see Figs. 1a and 1b), with minor distortions (Carugo, 2003; Berkholz *et al.*, 2012).

In contrast, the other two covalent bonds of the polypeptide backbone are much more interesting (see Fig. 1c). The rotation around the N–C<sup>α</sup> bond can be monitored by the dihedral angle C<sub>*i*-1</sub>–N–C<sup>α</sup>–C, which is named  $\varphi$ , while the rotation around the C<sup>α</sup>–C bond can be measured by the dihedral angle N–C<sup>α</sup>–C–N<sub>*i*+1</sub>, which is termed  $\psi$  (where C<sub>*i*-1</sub> and N<sub>*i*+1</sub> indicate the carbonylic C atom of the preceding residue and the amidic N atom of the following residue, respectively). Both of these covalent bonds are single and as a consequence there is the possibility of modifying the conformation of the molecule by rotating around them. However, the rotations are not completely free because of interatomic clashes that can occur during the rotation. Using a hard-sphere atomic model grounded on basic quantum-mechanics principles, atomic co-penetrations are impossible, *i.e.* ‘forbidden’.

G. N. Ramachandran and coworkers ideated a simple yet surprisingly efficient method to explore the energy landscape associated with this rotation using a small model compound: *N*-acetyl-L-alanine-methylamide (see Fig. 2). The chemical groups conjugated to the N- and the C-termini of alanine mimicked two residues, one preceding and the other following alanine. All possible  $\varphi$  and  $\psi$  value combinations were computationally generated and for each it was verified whether interatomic clashes occurred, assuming a hard-sphere atomic model. The  $\varphi/\psi$  space that can be populated by a peptide is only about one quarter of the theoretically available space.

An interatomic clash is unacceptable since atoms cannot co-penetrate each other. A collision occurs when the distance between two atoms is smaller than the sum of their van der Waals radii. G. N. Ramachandran and coworkers used two sets of van der Waals radii and it was thus possible to distinguish three cases: (i) two atoms are sufficiently distant, (ii) two atoms are moderately colliding (their distance is smaller than the sum of the largest van der Waals radii and larger than the sum of the smallest radii) and (iii) two atoms are colliding acutely (their distance is smaller than the sum of the smallest radii).

It was then possible to draw a simple bi-dimensional plot with the  $\varphi$  values on the horizontal axis and the  $\psi$  values on the vertical axis, and to divide it into three different zones: those where there were no interatomic clashes, those where there were moderate clashes and those where clashes were



**Figure 1**  
Polypeptide backbone and possible rotations. (a) Rotation around the peptide C–N bond is not possible because of the mesomeric equilibrium between two resonance structures. (b) As a consequence of the dihedral angle  $\omega$  may assume two values, yielding two conformations: *cis* ( $\omega = 0^\circ$ ) and *trans* ( $\omega = 180^\circ$ ). (c) The rotation around the N–C $^\alpha$  bond can be monitored by the dihedral angle C $_{i-1}$ –N–C $^\alpha$ –C, which is named  $\varphi$ , while the rotation around the C $^\alpha$ –C bond can be measured by the dihedral angle N–C $^\alpha$ –C–N $_{i+1}$ , which is termed  $\psi$ .

extremely severe (see Fig. 3). These three types of regions were named ‘fully allowed’, ‘partially allowed’ and ‘forbidden’. This bi-dimensional plot is the Ramachandran plot.

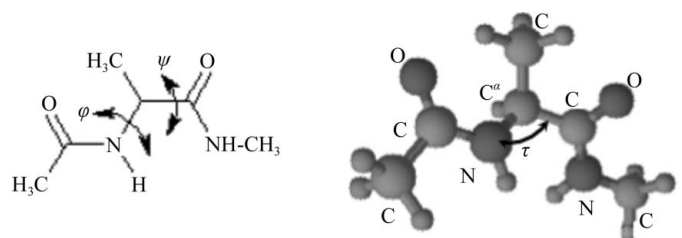
Two Ramachandran plots are shown in Fig. 3(b): one for one of the oldest protein crystal structures, dogfish M4 lactate dehydrogenase (PDB entry 3ldh; White *et al.*, 1976), and the other for a recent structure, *Escherichia coli* RNA chaperone Hfq in complex with ATP (PDB entry 3qo3; Hämmerle *et al.*, 2012). A more pronounced clustering within the ‘fully allowed’ region is evident in the plot for the more recent structure, a sign of the technological improvements in protein crystal structure determination that have occurred over the last four decades. Notably, the structure of M4 lactate dehydrogenase deposited in 1976 is at low resolution (3 Å) despite its relatively moderate size (330 amino acids) and it was only partially refined.

Additionally, it must be remembered that the original plots were only drawn for amino acids other than glycine and proline. These two residues are either too flexible (glycine, because of the absence of the side chain) or too rigid (proline, because of the presence of a penta-atomic heterocyclic ring) and their allowed stereochemistries differ from those of the other 18 L-amino acids. The Ramachandran plots for glycine and proline are shown in Fig. 4.

It is necessary to remember that there is a marked dependence of the Ramachandran plot on the bond angle N–C $^\alpha$ –C named  $\tau$  (see Fig. 2). For a regular *sp*<sup>3</sup> C atom, the value of  $\tau$  should be 109.5°. In proteins,  $\tau$  averages at 110°, although values as small as 100° and as large as 120° can be observed. Usually, Ramachandran plots for  $\tau = 110^\circ$  are used. However, the plots for different values of  $\tau$  are quite different.

## 2. Redrawing

The Ramachandran plot has repeatedly been reconsidered during its first half century of life (Bansal & Srinivasan, 2013) and especially during the last two decades, during which large numbers of three-dimensional structures of proteins have been determined and made available through the Protein Data Bank. In fact, while the original map was only based on theoretical computations, in more modern times Ramachandran plots are generally generated on the basis of experimental observations.



**Figure 2**  
The simple model compound *N*-acetyl-L-alanine-methylamide used by Ramachandran and coworkers to explore the conformational space defined by the two torsions  $\varphi$  and  $\psi$ . The bond angle defined by the N, C $^\alpha$  and C atoms is named  $\tau$ .

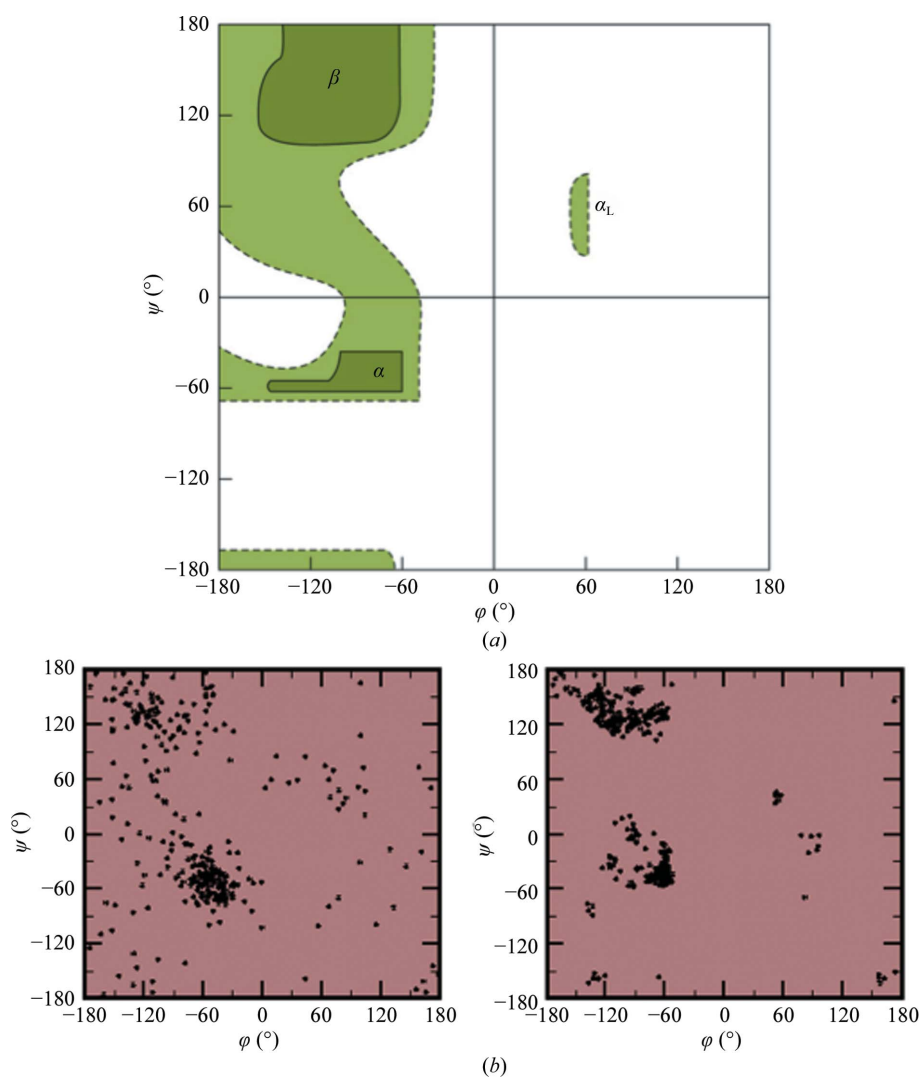
It is possible to observe several differences amongst different studies, where three main points emerge: (i) the amount of data used to redraw the Ramachandran plot, (ii) the control of the data quality and (iii) the criteria adopted to calculate tendencies and propensities from the maps. The amount of data is strictly dependent on the growth of the Protein Data Bank and it is obvious that more data were used in more recent studies. The quality of the data is also related to the size of the Protein Data Bank, in the sense that stricter criteria were used in more recent analyses when the resulting 'clean' data were sufficiently numerous. Different quality criteria have been used based on the crystallographic resolution, the  $R$  factor and the free  $R$  factor, the atomic displacement parameters and even the electron density (Giacovazzo *et al.*, 2011). Different criteria for transforming the Ramachandran plot, which is essentially a scatter plot, into a continuous conformational surface have also been used, ranging from

simple smoothing functions (Walther & Cohen, 1999) to more sophisticated kernel functions (Amir *et al.*, 2008), Fourier series (Pertsemliadis *et al.*, 2005) and Dirichlet models (Lovell *et al.*, 2003).

Without pretention to be exhaustive, some of the main contributions to the redrawing of the Ramachandran plot are briefly summarized below.

The interplay between side-chain conformation and backbone secondary structure was described early on (Dunbrack & Karplus, 1993; Schrauber *et al.*, 1993). By investigating the influence of the side-chain conformation on the accessible region of Ramachandran space, Chakrabarti and Pal later observed that for each stable side-chain conformation the residues reside in only a limited section of the allowed region of the Ramachandran plot (Chakrabarti & Pal, 1998). Although this study was limited to a small set of 120 protein structures available in 1998 and although the side-chain conformation was approximated as *trans* ( $\chi_1 \simeq 180^\circ$ ), *gauche*<sup>-</sup> ( $\chi_1 \simeq 60^\circ$ ) or *gauche*<sup>+</sup> ( $\chi_1 \simeq -60^\circ$ ) without considering a more realistic description (the real value of  $\chi_1$ ), it is not really unexpected that the backbone shape is also influenced by the steric needs and constraints of the side chains. This was ignored in the original analyses of the main-chain stereochemical features, which were performed on an alanine-like small model compound that lacked side-chain atoms beyond the  $C^\beta$  atom.

Moreover, several groups observed that the  $\varphi/\psi$  distributions tend to be narrower than previously predicted (Herzberg & Moulton, 1991; Jones & Thirup, 1986; Karplus, 1996). For example, using a few tens of protein structures, Herzberg and Moulton observed in 1991 that there are relatively few sterically strained main-chain dihedral angles and that distortions are overwhelmingly located in regions concerned with function (Herzberg & Moulton, 1991). Later, in 1999, Walther and Cohen observed substantially narrowed  $\varphi/\psi$  distributions at higher crystallographic resolutions by surveying 808 protein crystal structures (209 367 amino acids; Walther & Cohen, 1999). While at lower resolution there is a considerable tolerance in the  $\varphi$  and  $\psi$  values associated with each type of secondary structure, at higher resolution the regions of the Ramachandran plot that are populated become narrower. Similar results were also reported by Hovmöller and coworkers in 2002, who extended the analysis to



**Figure 3**

(a) Ramachandran plot for Ala and Ala-like residues, showing the allowed regions (continuous lines) and the partially allowed regions (dotted lines) (adapted from Ramakrishnan, 2001). (b) Ramachandran plots for dogfish M4 lactate dehydrogenase (PDB entry 3ldh; left) and for *E. coli* RNA chaperone Hfq in complex with ATP (PDB entry 3qo3; right).

each single type of amino acid by examining 1042 protein structural domains (237 384 amino acids; Hovmöller *et al.*, 2002). A subsequent version of the Ramachandran plot was generated in 2005 by Anderson and coworkers by using a larger data set of 4383 protein crystal structures and carefully scrutinizing their quality (Anderson *et al.*, 2005).

Several scientists also focused their attention on the dependence of the Ramachandran plot on the sequence neighbours of the residues.

In the 1990s, it was observed that the residues that precede proline often populate the region at  $\varphi = -130^\circ$  and  $\psi = +80^\circ$  (usually named the  $\zeta$  region; Karplus, 1996) and have a minor tendency to be observed in the  $\alpha$  region of the Ramachandran plot (MacArthur & Thornton, 1991).

Ting and coworkers examined the Ramachandran plots of each residue type as a function of the preceding or following residues in loops, thus excluding helices and strands (Ting *et al.*, 2010). In agreement with the results previously reported by Jha *et al.* (2005), it was observed that the propensity to adopt a certain backbone conformation is markedly influenced by the conformation of the neighbouring residues, indicating that the conformational space is considerably smaller than that predicted on the basis of the hypothesis that the conformation of each residue is independent of the shape of the rest of the protein.

The correlation between backbone conformation and amino-acid sequence was also shown by a statistical survey of tripeptides found in the Protein Data Bank (Keskin *et al.*, 2004). Furthermore, exhaustive computational enumeration of the allowed conformations in short polyalanines (acetyl-Ala<sub>n</sub>-N-methylamides;  $n = 6-9$ ) supports the hypothesis that the backbone stereochemistry of a residue is influenced by the geometry of the surrounding amino acids (Pappu *et al.*, 2000).

Several studies were also devoted to the special cases of glycine, proline and pre-proline residues. Glycines are much more flexible than the other residues since they lack a bulky side chain and can therefore assume conformations that are not possible for other residues. Their  $\varphi$ - $\psi$  maps have been investigated by both statistical surveys of experimental data

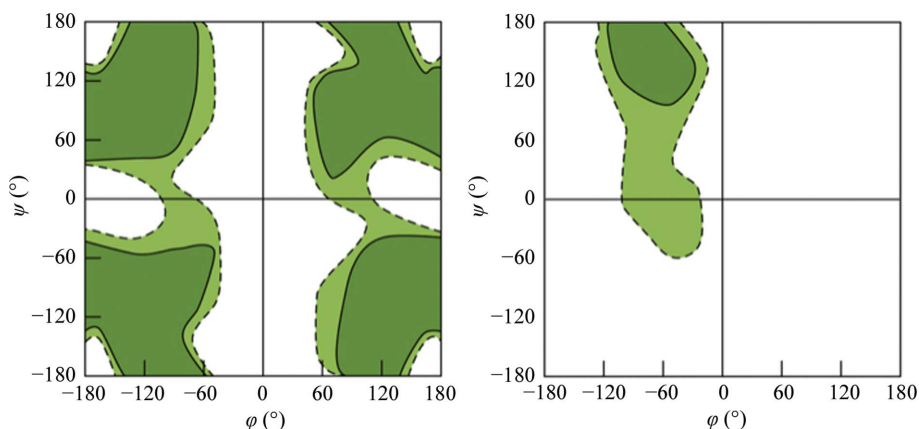
and by computational conformational searches (Ramakrishnan & Ramachandran, 1965; Hovmöller *et al.*, 2002; Hu *et al.*, 2003; Ho & Brasseur, 2005). In contrast, prolines are considerably more rigid than the other residues because of their side chain, which is conjugated to the amido N atom with the formation of a penta-atomic heterocyclic ring. The Ramachandran plot of proline has been investigated in detail (Summers & Karplus, 1990; Ho *et al.*, 2005), as well as that of the residues that precede a proline residue, the flexibility of which is markedly influenced by the presence of the proline (Ho & Brasseur, 2005; Summers & Karplus, 1990; MacArthur & Thornton, 1991; Karplus, 1996; Schimmel & Flory, 1968; Hurley *et al.*, 1992).

Eventually, after so many updates of the old Ramachandran concept, it turned out that the nomenclature associated with the various regions of the map which are populated by amino acids is increasingly complicated, irregular and varied. Hollingsworth and Karplus recently proposed a new nomenclature which might be adopted by the structural bioinformatics and structural biology communities (see Fig. 5; Hollingsworth & Karplus, 2010).

The  $\beta$  region, which occupies a large fraction of the north-western quadrant of the original Ramachandran plot, is now divided into two separate zones: one with residues that are actually found in  $\beta$ -strands ( $\beta$  region) and the other with residues that form polyproline II spirals (Woody, 2009) characterized by the absence of hydrogen bonds between the N-H group of a residue and the C=O group of one of the following residues ( $P_{II}$  region). A third region, which is very narrow and centred around  $\varphi = -63^\circ$  and  $\psi = -43^\circ$ , is occupied by residues that form  $\alpha$ -helices (and is thus named  $\alpha$ ). A region close to the latter and expanding in the direction of zone  $\beta$ , which is often referred to as the bridge sector, is proposed to be the  $\delta$  region. A fifth zone, named  $\gamma$ , is allowed for residues that form  $\gamma$  turns, which have an  $O_i \cdots NH_{i+2}$  hydrogen bond (Némethy & Printz, 1972; Matthews, 1972). Additionally, there are the  $\zeta$  region, which is often populated by residues that precede proline (but also by other residues), and the  $\varepsilon$  region, which is sparsely populated, mostly by glycines, at positive  $\varphi$  values in the north-eastern and south-eastern quadrants of the Ramachandran plot, respectively.

All of these regions are named with the initial letters of the Greek alphabet ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$  and  $\zeta$ ), with the exception of the  $P_{II}$  region, the naming of which has historical reasons and is deeply rooted in the protein structure lexicon.

Moreover, in addition to these seven regions that can be populated by protein residues there are their mirror images. In reality, these are not mirror images in stereochemical terms (they are impossible given the chiral nature of the natural amino acids). However, they are usually named mirror images since



**Figure 4** Ramachandran plots for glycine (left) and proline (right), showing the the allowed regions (continuous lines) and the partially allowed regions (dotted lines) (adapted from Ramakrishnan, 2001).

ribbons approximate the backbone conformation. As a consequence, it is possible to define mirror images for the  $\alpha$ ,  $\gamma$ ,  $\delta$  and  $P_{II}$  regions and these regions are named by adding a prime ( $\alpha'$ ,  $\gamma'$ ,  $\delta'$  and  $P'_{II}$ ). For example, left-handed  $\alpha$ -helices are the mirror images of the more common right-handed  $\alpha$ -helices.

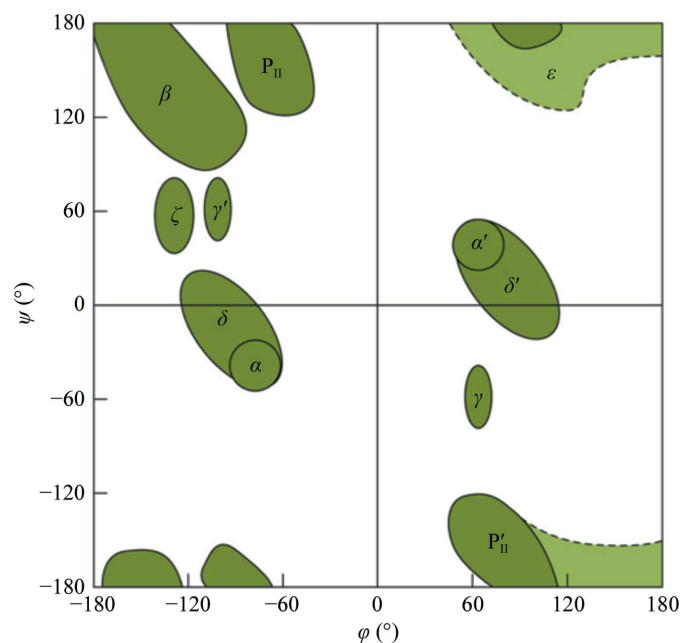
In contrast, it is impossible to define a mirror image of the  $\beta$  and the  $\varepsilon$  regions, since they would be nearly indistinguishable from the  $\beta$  and  $\varepsilon$  regions themselves, and of the  $\zeta$  region, since the  $C^\beta$  atoms would collide severely.

With the exception of the  $\gamma'$  region, which is much more populated than the  $\gamma$  region, all of the other mirror-image zones are little populated. In particular, very few residues are observed in the  $\alpha'$  zone of the Ramachandran plot (Hollingsworth & Karplus, 2010).

### 3. Keep out: disallowed

Early in the history of protein structures, it became apparent that regions of the Ramachandran plot that are in principle disallowed are nevertheless sporadically populated. This apparent violation of the stereochemical rules defined fifty years ago by G. N. Ramachandran has been investigated several times and it must be observed that many regions that were originally considered to be prohibitively unfavourable were later discovered to be allowed and characteristic of several types of stable backbone conformations such as, for example,  $\gamma$  and  $\beta$  turns.

It was estimated that only 0.3% of residues are observed in disallowed zones of the Ramachandran plot (Gunasekaran *et al.*, 1996). Although this type of evaluation depends on the



**Figure 5**

The nomenclature proposed by Hollingsworth & Karplus (2010). Note that the region for left-handed helices ( $\alpha'$ ) was not considered by Hollingsworth & Karplus (2010) since it is basically empty. This is a qualitative depiction freely inspired by the original publication.

exact definition of what is allowed and disallowed and on the data set of protein structures that are taken into consideration, similar estimations were reported for both peptide (1.0%; Ramakrishnan *et al.*, 2007) and protein structures [0.4% (Pal & Chakrabarti, 2002) and 0.6% (Ramakrishnan *et al.*, 2007)].

Most of the amino-acid residues with unfavourable  $\varphi/\psi$  torsions are found in loops and irregular structures or at the beginning or end of helices and strands (Gunasekaran *et al.*, 1996). In the analysis by Gunasekaran *et al.* (1996) one third of the conformationally anomalous amino-acid residues occur in long loops, while in the study by Pal & Chakrabarti (2002) they are often observed in short loops. In this study it was observed that they tend to be solvent-exposed, while in the aforementioned analysis it appears that there is no preference for high or low solvent accessibility (Gunasekaran *et al.*, 1996).

Anomalous backbone conformations were mainly rationalized in two alternative, and not mutually exclusive, ways. On the one hand, it was proposed that distortions of the  $\varphi/\psi$  angles away from the allowed regions of the Ramachandran plot were compensated by other interactions (such as, for example, hydrogen bonds, metal-cation coordination or dipole–dipole interactions) that occurred in the surroundings of the residues that were observed in disallowed zones of the Ramachandran plot. On the other hand, it was observed that distortions in bond lengths and angles might relieve the local strain in the  $\varphi/\psi$  torsions.

For example, Deane *et al.* (1999) observed that the unfavourable  $\varphi$  and  $\psi$  combinations that are often observed for Asp and Asn residues can be compensated by favourable interactions between the dipoles associated with the carbonyls. The dipole associated with the Asn/Asp side-chain C=O can be attracted by the dipole associated with the main-chain C=O of the preceding residue or of the Asn/Asp itself. Several studies suggest that the energy associated with these nonbonding dipole–dipole interactions can be of the same order of magnitude as hydrogen bonds (Maccallum *et al.*, 1995*a,b*; Allen *et al.*, 1998). Vega and colleagues, moreover, showed that residue Asn47 of the  $\alpha$ -spectrin SH3 domain can assume unfavourable  $\varphi/\psi$  torsions in a type II'  $\beta$ -turn and that its mutation to glycine has a modest impact on structure, folding stability and folding kinetics (Vega *et al.*, 2000).

In 1996, Karplus and Gunasekaran and coworkers observed that unfavourable  $\varphi/\psi$  torsions are compensated in many cases by distortions of the bond angles centred on the  $C^\alpha$  atom (Karplus, 1996; Gunasekaran *et al.*, 1996). This was later confirmed by Ramakrishnan *et al.* (2007). In other words, some regions that are considered to be extremely unfavourable can be accessed if some covalent bonds slightly lengthen and if some bond angles widen/tighten. In particular, Ramakrishnan and coworkers observed that 'none of the 88 examples of disallowed conformations observed in peptide and protein structures is accompanied by convincing short contacts in the crystal structures' (Ramakrishnan *et al.*, 2007). However, Karplus also observed that residues which occupy the disallowed region that links the  $\alpha$  and the  $\beta$  zones can be stabilized by electrostatic interactions between the N–H<sub>*i*+1</sub> and the N<sub>*i*</sub> groups, which are made possible by some

distortions of the bond angles centred on the C<sup>α</sup> atom (Karplus, 1996).

Recently, Porter and Rose focused attention on a zone of the Ramachandran plot that is usually not very populated: the 'bridge region' defined by  $\varphi < 0^\circ$  and  $-20 \leq \psi \leq 40^\circ$  (Porter & Rose, 2011a). According to their data, if a residue adopts a backbone conformation that corresponds to the  $\varphi/\psi$  combinations in the bridge region, the amide N atom of the next residue cannot form hydrogen bonds to water molecules. A residue can therefore fall within the bridge region which would become sterically allowed only if the amide N atom of the following residue forms intrapeptide hydrogen bonds in the folded protein. This implies an expansion of the space that is accessible to protein conformations.

A different hypothesis was proposed by Regan and coworkers, who showed that the fraction of residues with  $\varphi/\psi$  combinations in the bridge region of the Ramachandran plot increases if the  $\tau$  bond angle centred on the C<sup>α</sup> atom (N–C<sup>α</sup>–C; see Fig. 2) widens (Porter & Rose, 2011b; Zhou *et al.*, 2011a,b). This fits perfectly with the earlier predictions of Ramachandran made fifty years ago and does not require invoking hydrogen bonds to explain the backbone conformations of proteins as in the hypothesis of Porter & Rose (2011a).

The discrepancy between the theses of Porter and Rose on the one hand and of Regan and coworkers on the other has been publicly commented on (Porter & Rose, 2011b; Zhou *et al.*, 2011a,b). However, it must be observed that the two theses are in reality not mutually exclusive and that further statistical surveys might be necessary to clarify this divergence.

#### 4. Applications

Ramachandran plots have been used during the last two decades to validate protein three-dimensional structures determined using crystallographic methods, NMR spectroscopy or even computational modelling techniques. The essential idea is the following. A residue with anomalous  $\varphi$  and  $\psi$  torsions, which are far from the allowed region of the Ramachandran plot, can be suspected to be wrong. In other words, if the backbone stereochemistry is far from what is expected, it can be hypothesized that a local mistake has occurred in the determination of the position of the atoms. It was observed that since the torsions  $\varphi$  and  $\psi$  are usually not restrained during refinement, the Ramachandran plot is a powerful validation tool (Kleywegt & Jones, 1996).

The first application of Ramachandran plots to the problem of protein structure validation was the software suite *PROCHECK* (Laskowski *et al.*, 1993), in which the Ramachandran space was divided into four regions (most favoured, additionally allowed, generously allowed and disallowed) according to a previous statistical survey of protein structures (Morris *et al.*, 1992). It was proposed that a 'good' structure should have more than 90% of its residues in the most favoured regions. Other applications were published later. *PROCHECK* was modified to also take into account solution structures determined by NMR spectroscopy (Laskowski *et*

*al.*, 1996). A *PROCHECK*-like approach was implemented into *WHAT\_CHECK* (Hooft & Vriend, 1997) and differently defined Ramachandran plots are also used in *MolProbability* (Chen *et al.*, 2010). A systematic comparison of several different validation tools was published in 1998 (EU 3-D Validation Network, 1998) and a survey of numerous validation tools was published in 2011 (Read *et al.*, 2011).

It is important to mention at this point that to date it has been very common to continue to divide the Ramachandran  $\varphi/\psi$  space into discrete regions of different 'quality' that are, in the case of *PROCHECK* for example, called most favoured, additionally allowed, generously allowed and disallowed. This is obviously a rather crude approximation since the conformational energy function is continuous, and more sophisticated alternatives are possible in which the position of a residue on the map may be associated with a real energy value.

A Ramachandran-like approach was published by Sims and Kin, who considered polypeptide fragments, resulting in vectors of two or more  $\varphi$  and  $\psi$  torsions and polydimensional conformational maps (Sims & Kim, 2006). In addition to the  $\varphi$  and  $\psi$  torsions, Tosatto and Battistuta also analyzed the side-chain dihedral angles and defined a 'conformational status' of a residue on the basis of all of its torsions (Tosatto & Battistuta, 2007). The TAP score of a protein structure, based on the conformational statuses of the residues, was able to effectively evaluate subtle distortions from the native protein structure (Tosatto & Battistuta, 2007).

Besides structure validation, Ramachandran plots have been used for many other purposes. For example, Dahl and coworkers published a classification of the amino acids based on their distribution on the Ramachandran plane, with unexpected features such as cysteine being grouped together with some aromatic residues (Tyr, Phe and His) and tryptophan being clustered with threonine (Dahl *et al.*, 2008).

Computational methods that include  $\varphi$ - $\psi$  restraints to help protein solution structure determination by NMR spectroscopy have also been designed and used (Bertini *et al.*, 2003; Kuszewski *et al.*, 1997, 1996).

Obviously, Ramachandran plots have also been used to define secondary structures and to assign them to amino-acid residues on the basis of their three-dimensional structures (Venkata *et al.*, 2010; Muñoz & Serrano, 1994; Kolaskar & Sawant, 1996; Gromiha *et al.*, 2002). This procedure has an advantage over several alternatives in that it does not require additional parameters such as those that are necessary, for example, to define hydrogen bonds.

Ramachandran plots have also been used to characterize amino acids since it is well known that different amino acids have different Ramachandran plots (Karplus, 1996; Hollingsworth & Karplus, 2010; Hovmöller *et al.*, 2002; Beck *et al.*, 2008; Berkholz *et al.*, 2009; Dahl *et al.*, 2008) and also different  $\varphi$  intrinsic propensities (Serrano, 1995). However, it has been proposed (Berkholz *et al.*, 2009; Hollingsworth & Karplus, 2010) that these maps do not indicate the energetic preferences of the residues for one or other region of the map. On the contrary, the differences in the maps indicate the different preferences for certain regions amongst different residues. For

example, the region close to  $\varphi = 90^\circ$  and  $\psi = 0^\circ$  ( $\delta'$  zone) can essentially be populated only by Gly, while the region close to  $\varphi = -90^\circ$  and  $\psi = 0^\circ$  ( $\delta$  zone) is accessible to nearly all nonproline residues. Therefore, the Ramachandran plot of Gly shows many points close to  $\varphi = 90^\circ$  and  $\psi = 0^\circ$  and fewer points close to  $\varphi = -90^\circ$  and  $\psi = 0^\circ$ . In fact, Gly has no competitors in the first region and 18 competitors in the second region. This hypothesis is strictly based on numerous observations that many mutations do not substantially change the fold stereochemistry and that mutations do not substantially change the  $\varphi$  and  $\psi$  torsions.

Another recent application of the Ramachandran plot is the development of new strategies to compare three-dimensional structures of proteins. The main idea is to represent the structure by a linear string of characters in such a way that the comparison of two structures becomes a simple comparison between two words/sequences, which is possible by using one of the numerous techniques that have been developed to align protein/DNA sequences and scan sequence databases (Kirilova & Carugo, 2008; Carugo, 2006, 2007). In the method published by Lo and coworkers, each amino-acid residue is associated with a letter of the alphabet as a function of its position in the Ramachandran plot. A modified version of the *BLAST* program (Altschul *et al.*, 1990) is then used to scan larger structure databases. Each comparison of two structures is impressively fast (about  $10^{-5}$  s on a 3.2 GHz CPU): nearly 250 000 faster than a widely used computer program such as *CE* (Shindyalov & Bourne, 1998). It is important to emphasize that Ramachandran plot analyses have also been incorporated in molecular-mechanics force fields and software (Brooks *et al.*, 2009; Fleishman *et al.*, 2010).

## 5. Extensions

An extension of the concept that supports the Ramachandran plot has recently been proposed. While each residue of a protein is represented by a point on the Ramachandran plot, each protein of an ensemble of proteins is represented by a point on the proteomic Ramachandran plot (PRplot; Carugo & Djinić-Carugo, 2013).

This is achieved by computing the circular average of the  $\varphi$  and  $\psi$  dihedral angles for each protein and by plotting the corresponding point on the map. By using a nonredundant set of protein structures taken from the PDBselect database (Griep & Hobohm, 2010), it was possible to verify that proteins are distributed around a sigmoid function such as, for example,

$$\psi = -165.7 \exp[-(88.7/x)^{-14}] + 126.9, \quad (1)$$

with a correlation coefficient equal to 0.936 (see Fig. 6). Closely similar expressions were obtained by using other nonredundant sets of proteins obtained from the SCOP (Andreeva *et al.*, 2008) database of protein structural domains or built using the *PISCES* web server (Wang & Dunbrack, 2003), or by using other sigmoid functions. Although the sigmoid correlation between  $\varphi$  and  $\psi$  lacks a specific physical explanation, it is possible to imagine it as the trajectory along which a protein structure can move on a  $\varphi/\psi$  map, although

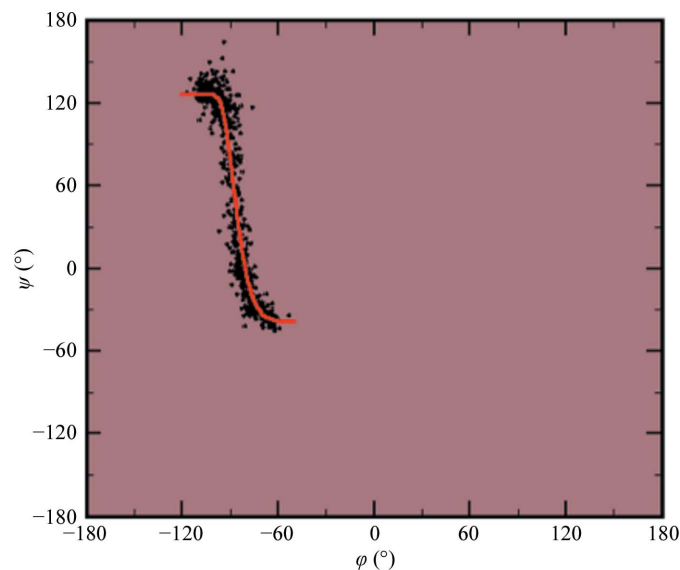
this does not mean that strands are physically replaced by helices (or *vice versa*) along this curve during evolution. However, it is clear that proteins can only occupy a very limited region of the PRplot and tend to cluster along and around the sigmoid curve.

## 6. Ramachandran plots and the internet

Although a Ramachandran plot can be produced manually, it is faster to make it with a computer and many programs offer this possibility. Here, we mention only web-based servers and applications that are freely available. One of them (<http://dicsoft1.physics.iisc.ernet.in/rp/>) allows numerous possibilities to analyze specific features of the Ramachandran plot (different types of amino acids, sequence-dependent windows, different regions of the map *etc.*; Sheik *et al.*, 2002; Gopalakrishnan, Sowmiya *et al.*, 2007). *RAMPAGE* is another well designed program from the *CCP4* software suite (Winn *et al.*, 2011) that can be found at <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>, for example.

Alternatively, there are databases focused on torsions (both main and side chains), such as, for example, *CADB-3.0* (<http://cluster.physics.iisc.ernet.in/cadb>; Gopalakrishnan, Sheik *et al.*, 2007), that allow detailed inspections of many aspects of torsional space. Another server that allows the production of Ramachandran plots for structures deposited in the Protein Data Bank is available at <http://eds.bmc.uu.se/ramachan.html> (Kleywegt & Jones, 1996).

Notably, many validation programs freely available on the web allow one to produce high-quality Ramachandran plots, such as, for example, *MolProbity* (Chen *et al.*, 2010). Additionally, many computer-graphics programs allow the production of Ramachandran plots, such as, for example, *PyMOL* (<http://www.pymol.org/>).



**Figure 6** Proteomic Ramachandran plot (PRplot). Each protein is represented by a point (black stars) and the ensemble of proteins is distributed along a sigmoid function (red curve).

Both of the authors are the same age as the Ramachandran plot and are prone to believe that this is a happy circumstance.

## References

- Allen, F. H., Baalham, C. A., Lommerse, J. P. M. & Raithby, P. R. (1998). *Acta Cryst.* **B54**, 320–329.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
- Amir, E. D., Kalisma, N. & Keasar, C. (2008). *Proteins*, **72**, 62–73.
- Anderson, R. J., Weng, Z., Campbell, R. K. & Jiang, X. (2005). *Proteins*, **60**, 679–689.
- Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2008). *Nucleic Acids Res.* **36**, D419–D425.
- Bansal, M. & Srinivasan, N. (2013). Editors. *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*. Singapore: World Scientific.
- Beck, D. A., Alonso, D. O., Inoyama, D. & Daggett, V. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 12259–12264.
- Berkholz, D. S., Driggers, C. M., Shapovalov, M. V., Dunbrack, R. L. & Karplus, P. A. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 449–453.
- Berkholz, D. S., Shapovalov, M. V., Dunbrack, R. L. & Karplus, P. A. (2009). *Structure*, **17**, 1316–1325.
- Bertini, L., Cavallaro, G., Luchinat, C. & Poli, I. (2003). *J. Biomol. NMR*, **26**, 355–366.
- Brooks, B. R. *et al.* (2009). *J. Comput. Chem.* **30**, 1545–1614.
- Carugo, O. (2003). *Acta Chim. Slov.* **50**, 505–511.
- Carugo, O. (2006). *Curr. Bioinformatics*, **1**, 75–83.
- Carugo, O. (2007). *Curr. Protein Pept. Sci.* **8**, 219–241.
- Carugo, O. & Djinović-Carugo, K. (2013). *Amino Acids*, **44**, 781–790.
- Chakrabarti, P. & Pal, D. (1998). *Protein Eng.* **11**, 631–647.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Dahl, D. B., Bohanan, Z., Mo, Q., Vannucci, M. & Tsai, J. (2008). *J. Mol. Biol.* **378**, 749–758.
- Deane, C. M., Allen, F. H., Taylor, R. & Blundell, T. L. (1999). *Protein Eng.* **12**, 1025–1028.
- Dunbrack, R. L. & Karplus, M. (1993). *J. Mol. Biol.* **230**, 543–574.
- EU 3-D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.
- Fleishman, S. J., Corn, J. E., Strauch, E. M., Whitehead, T. A., Andre, I., Thompson, J., Havranek, J. J., Das, R., Bradley, P. & Baker, D. (2010). *Proteins*, **78**, 3212–3218.
- Giacovazzo, C., Monaco, H. L., Artioli, G. & Viterbo, D. (2011). *Fundamentals of Crystallography*. Oxford University Press.
- Gopalakrishnan, K., Sheik, S. S., Vasuki Ranjani, C., Udayakumar, A. & Sekar, K. (2007). *Protein Pept. Lett.* **14**, 665–668.
- Gopalakrishnan, K., Sowmiya, G., Sheik, S. S. & Sekar, K. (2007). *Protein Pept. Lett.* **14**, 669–671.
- Griep, S. & Hobohm, U. (2010). *Nucleic Acids Res.* **38**, D318–D319.
- Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H. & Sarai, A. (2002). *Biopolymers*, **64**, 210–220.
- Gunasekaran, K., Ramakrishnan, C. & Balaram, P. (1996). *J. Mol. Biol.* **264**, 191–198.
- Hämmerle, H., Beich-Frandsen, M., Večerek, B., Rajkowitzsch, L., Carugo, O., Djinović-Carugo, K. & Bläsi, U. (2012). *PLoS One*, **7**, e50892.
- Herzberg, O. & Moulton, J. (1991). *Proteins*, **11**, 223–229.
- Ho, B. K. & Brasseur, R. (2005). *BMC Struct. Biol.* **5**, 14.
- Ho, B. K., Coutsiyas, E. A., Seok, C. & Dill, K. A. (2005). *Protein Sci.* **14**, 1011–1018.
- Hollingsworth, S. A. & Karplus, P. A. (2010). *Biomol. Concepts*, **1**, 271–283.
- Hoof, R. W. W. & Vriend, G. (1997). *Comput. Appl. Biosci.* **13**, 425–430.
- Hovmöller, S., Zhou, T. & Ohlson, T. (2002). *Acta Cryst.* **D58**, 768–776.
- Hu, H., Elstner, M. & Hermans, J. (2003). *Proteins*, **50**, 451–463.
- Hurley, J. H., Mason, D. A. & Matthews, B. W. (1992). *Biopolymers*, **32**, 1443–1446.
- Jha, A. K., Colubri, A., Zaman, M. H., Koide, S., Sosnick, T. R. & Freed, K. F. (2005). *Biochemistry*, **44**, 9691–9702.
- Jones, T. A. & Thirup, S. (1986). *EMBO J.* **5**, 819–822.
- Karplus, P. A. (1996). *Protein Sci.* **5**, 1406–1420.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *Nature (London)*, **181**, 662–666.
- Keskin, O., Yuret, D., Gursoy, A., Turkay, M. & Erman, B. (2004). *Proteins*, **55**, 992–998.
- Kirillova, S. & Carugo, O. (2008). *BMC Res. Notes*, **1**, 44.
- Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1385–1400.
- Kolaskar, A. S. & Sawant, S. (1996). *Int. J. Pept. Protein Res.* **47**, 110–116.
- Kuszewski, J., Gronenborn, A. M. & Clore, G. M. (1996). *Protein Sci.* **5**, 1067–1080.
- Kuszewski, J., Gronenborn, A. M. & Clore, G. M. (1997). *J. Magn. Reson.* **125**, 171–177.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). *J. Biomol. NMR*, **8**, 477–486.
- Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- MacArthur, M. W. & Thornton, J. M. (1991). *J. Mol. Biol.* **218**, 397–412.
- Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995a). *J. Mol. Biol.* **248**, 361–373.
- Maccallum, P. H., Poet, R. & Milner-White, E. J. (1995b). *J. Mol. Biol.* **248**, 374–384.
- Matthews, B. W. (1972). *Macromolecules*, **5**, 818–819.
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
- Muñoz, V. & Serrano, L. (1994). *Proteins*, **20**, 301–311.
- Némethy, G. & Printz, M. P. (1972). *Macromolecules*, **5**, 755–758.
- Pal, D. & Chakrabarti, P. (2002). *Biopolymers*, **63**, 195–206.
- Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 12565–12570.
- Pertsemliadis, A., Zelinka, J., Fondon, J. W. III, Henderson, R. K. & Otwinowski, Z. (2005). *Stat. Appl. Genet. Mol. Biol.* **4**, doi:10.2202/1544-6115.1165.
- Porter, L. L. & Rose, G. N. (2011a). *Proc. Natl Acad. Sci. USA*, **108**, 109–113.
- Porter, L. L. & Rose, G. N. (2011b). *Protein Sci.* **20**, 1771–1773.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
- Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
- Ramakrishnan, C. (2001). *Resonance*, **6**, 48–56.
- Ramakrishnan, C., Lakshmi, B., Kurien, A., Devipiya, D. & Srinivasan, N. (2007). *Protein Pept. Lett.* **14**, 672–682.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
- Read, R. J. *et al.* (2011). *Structure*, **19**, 1395–1412.
- Sasisekharan, V. (1962). *Collagen*, edited by N. Ramanathan, pp. 39–78. New York: John Wiley & Sons.
- Schimmel, P. R. & Flory, P. J. (1968). *J. Mol. Biol.* **34**, 105–120.
- Schrauber, H., Eisenhaber, F. & Argos, P. (1993). *J. Mol. Biol.* **230**, 592–612.
- Serrano, L. (1995). *J. Mol. Biol.* **254**, 322–333.
- Sheik, S. S., Sundararajan, P., Hussain, A. S. & Sekar, K. (2002). *Bioinformatics*, **18**, 1548–1549.
- Shindyalov, I. N. & Bourne, P. E. (1998). *Protein Eng.* **11**, 739–747.



- Sims, G. E. & Kim, S.-H. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4428–4432.
- Summers, N. L. & Karplus, M. (1990). *J. Mol. Biol.* **216**, 991–1016.
- Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M. I. & Dunbrack, R. L. (2010). *PLoS Comput. Biol.* **6**, e1000763.
- Tosatto, S. C. & Battistutta, R. (2007). *BMC Bioinformatics*, **8**, 155.
- Vega, M. C., Martínez, J. C. & Serrano, L. (2000). *Protein Sci.* **9**, 2322–2328.
- Venkata, M., Kumar, S. & Swaminathan, R. (2010). *Proteins*, **78**, 900–916.
- Walther, D. & Cohen, F. E. (1999). *Acta Cryst.* **D55**, 506–517.
- Wang, G. & Dunbrack, R. L. (2003). *Bioinformatics*, **19**, 1589–1591.
- White, J. L., Hackert, M. L., Buehner, M., Adams, M. J., Ford, G. C., Lentz, P. J., Smiley, I. E., Steindel, S. J. & Rossmann, M. G. (1976). *J. Mol. Biol.* **102**, 759–779.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Woody, R. W. (2009). *J. Am. Chem. Soc.* **131**, 8234–8245.
- Zhou, A. Q., O'Hern, C. S. & Regan, L. (2011*a*). *Protein Sci.* **20**, 1166–1171.
- Zhou, A. Q., O'Hern, C. S. & Regan, L. (2011*b*). *Protein Sci.* **20**, 1774.