The pre-proline Ramachandran plot

Schimmel and Flory argued in 1968 that pre-proline – amino acids preceding proline – has a particularly restricted Ramchandran plot, compared to the generic Ramachandran plot. This was finally observed in the protein database by MacArthur and Thornton (Figure 1B).

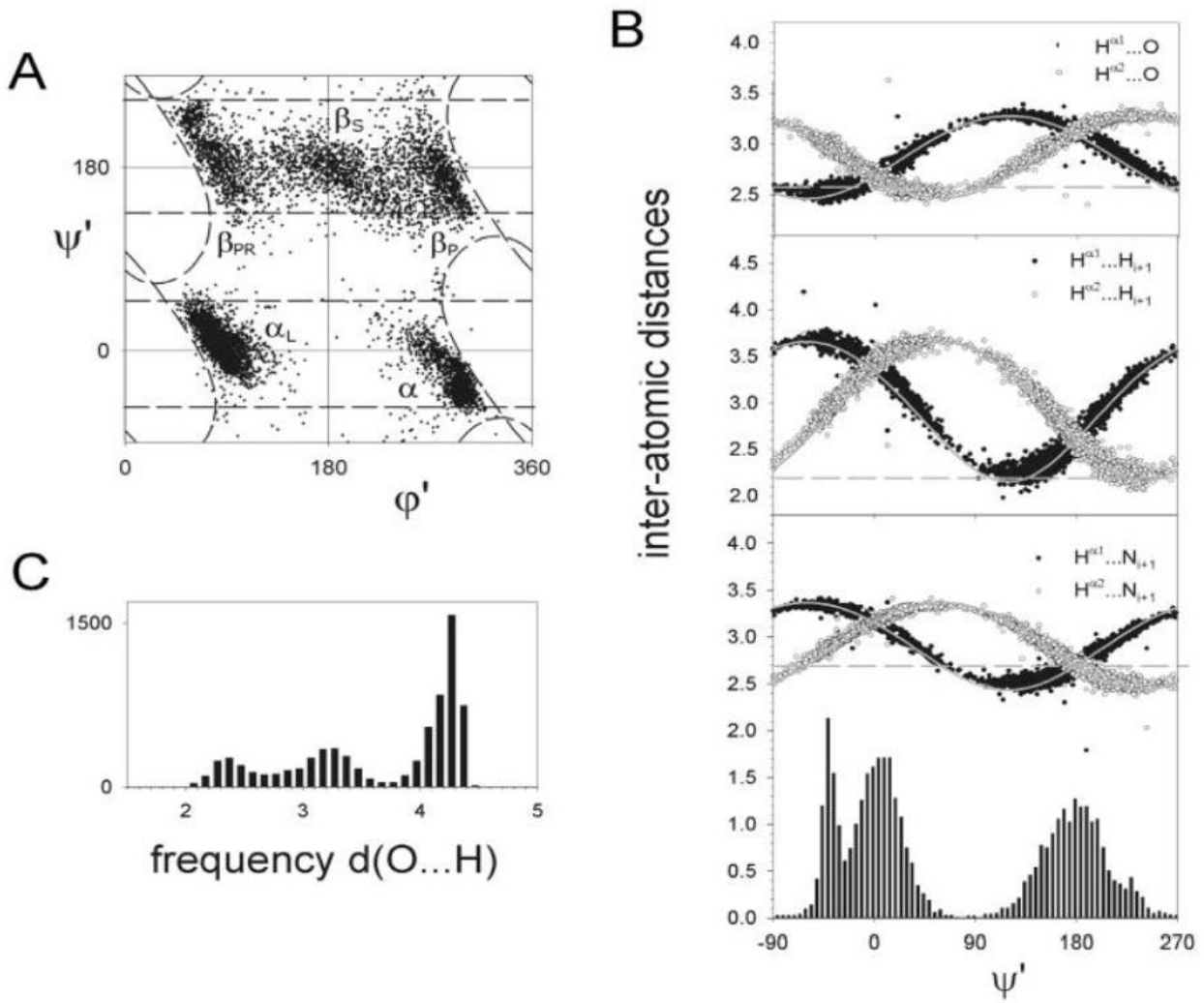

Figure 3
Glycine parameters. (A) The Ramachandran plot in shifted coordinates φ'-ψ'. The dashed lines show the steric clashes that define the boundaries of the observed densities (Figure 2B describes the specific interactions). (B) The distributions of various inter-atomic interactions as a function of ψ'. The dashed line show the limit of the VDW diameters. The grey line gives the model curve calculated with ideal geometry. At the bottom is the frequency distribution of the ψ' angle. (C) Frequency distribution of the inter-atomic distance d(O···H). There are 3 peaks, of which, the smallest at d(O···H) = 2.4 Å, which corresponds to the β_S region.

There are three main differences between the pre-proline Ramachandran plot and the generic Ramachandran plot. In the pre-proline Ramachandran plot, there is a large excluded horizontal strip at $-40° < \psi < 50°$, which restricts α_L and α regions. The α_L region is shifted up higher. These two features were reproduced in the Schimmel-Flory cal- culation and subsequent calculations. The third feature is a little leg of density poking out below the β-region (Figure 1B; purple in Figure 2C). Karplus called this the

ζ region, which is unique to pre-proline. Previous calculations did not focus on the indi- vidual interactions, and did not account for the ζ region. Here, we identify the exact steric clashes that determine the pre-proline Ramachandran plot. We will then analyse the interactions responsible for the ζ region.

Genesis for the Ramachandran Map

During the 50's and 60's, the focus of Ramachandran's work in the field of biophysics has been the elucidation of the structure of the fibrous protein collagen. He and his group proposed a modelled structure for collagen based on X-ray diffraction and related data. This came to be known as two-bonded structure, since it had two inter-chain hydrogen bonds per three residue-repeat. However, this structure was subjected to criticism by two British scientists, Rich and Crick on the basis of steric hin-drance. To answer this, a survey was made of the available crystal structures that showed that such short interatomic distances were existent in some structures. This observation gave rise to a new idea of using this information for checking any peptide/ protein conformation. Thus the seed was sown for a work of a more fundamental nature, which later took shape as Ramachandran map, capable of being used with protein struc-tures.

Basics for the Map

Parameter: The basic parameter that was used is known as the torsion angle (synonyms being dihedral angle and twist angle). The torsion angle is applicable to any part of a molecule, which

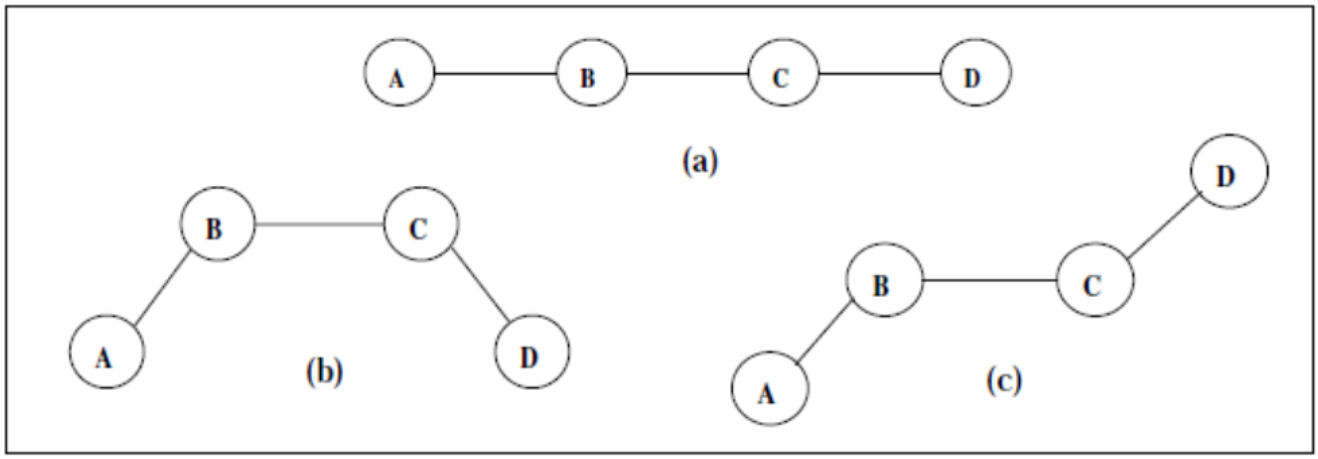

Figure 1. (a) Schematic representation of a system of 4 atoms A, B, C, D linked linearly by bonds. Orientation in the system ABCD for (b) $\sigma = 0°$. (c) $\sigma = 180°$. Note that in (b) atoms A and D are on the same side(cis) of B-C, and in (c) they are on the opposite sides (trans) of B-C.

has a system of four atoms A, B, C and D connected linearly through covalent bonds. This is shown schematically in *Figure* 1a. It is necessary to keep in mind that torsion angle is not just a geometrical angle (such as ABC etc.) but an angle of rotation. Hence it requires unambiguous definition about the starting position for measuring rotation and sense of rotation. The symbol employed to represent any torsion angle is σ (Greek letter chi) and two representative positions are shown in *Figure* 1b and 1c corresponding to $\sigma = 0°$ and $\sigma = 180°$. In chemistry literature, these are known as *cis* and *trans* orientations respectively (since atoms A and D are on the same side of BC in *Figure* 1b and on opposite sides in *Figure* 1c).

System used: The very first step is to choose a suitable system on which this concept of steric hindrance could be applied. As is well known in protein chemistry, any protein chain is made up of a large number of amino acids belonging to twenty different varieties, linked in a sequential way through amide/peptide linkages. An alternate way of looking at a protein chain is that it has a backbone chain of atoms with the amino acid side group atoms as branches at the α-carbon atoms. Keeping in mind that linkage of two amino acids gives rise to a geometrically well-defined peptide unit, the backbone of a protein with n amino acid residues will contain $n-1$ peptide units. The geometry of the peptide units (viz., bond lengths, bond angles) had been given as early as 1953 by Linus Pauling and his group and the crystal structures of the large number of proteins available nowadays

stand proof of the almost invariant nature of the parameters of the peptide unit. Such a peptide unit in *trans*-configuration spanning from one C^a to the next is shown in *Figure* 2 (as a ball and stick model drawn by a software package RASMOL). The basic system that was used in the formulation of Ramachandran map is a system of two such linked peptide units, shown in *Figure* 3.

Ramachandran angles **(f,y)**: Having defined a parameter (torsion angle) and chosen a system (two linked peptide units), the location of the parameter in the sys-tem had to be decided. The two bonds N – C^a and C^a– C are the ones about which rotations are permissible and hence torsion angles have been defined with respect to these bonds. They are

(i) σ (C – N – C^a – C) and (ii) σ (N – C^a – C – N).

These two torsion angles were assigned specific symbols, the first one as f(phi) and the second one as y (psi). These came to be known as Ramachandran angles. Every junction, therefore, would have a unique pair of Ramachandran angles (f, y) defining the relative orientations of the adjoining peptide units. It is easily seen that a protein chain containing n amino acids will have

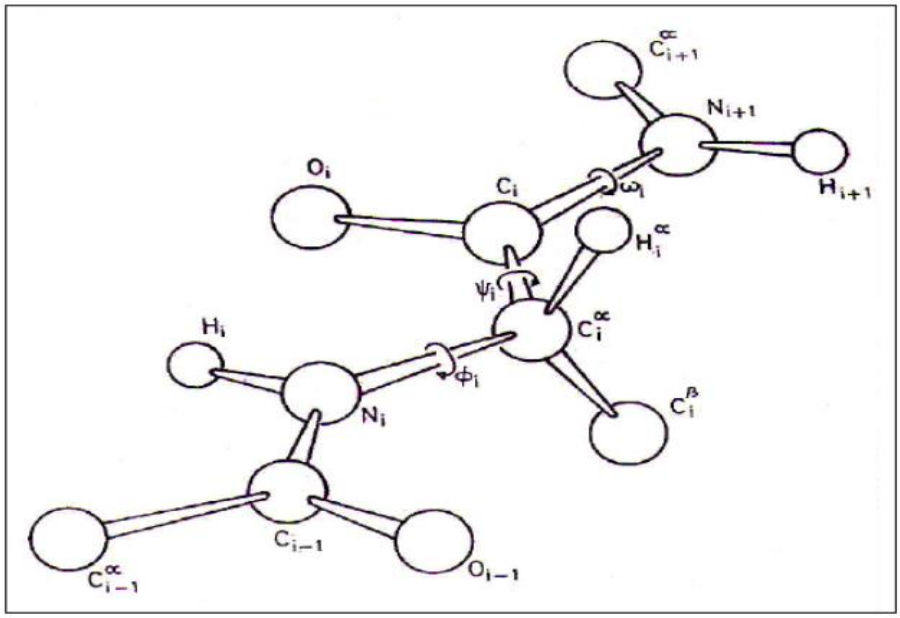

Figure 3. A system of two trans-planar peptide units linked at the central a-carbon atom. Ramachandran angles (f_i, y_i) are shown.

n-1 peptide units, n–2 junctions and the backbone of the protein chain can be described in terms of the n–2 pairs of Ramachandran angles (f, y), one at every junction.

Principle Underlying the Formation of Ramachandran Map

It had been believed that the atoms in a structure could not approach each other closer than the sum of their van der Waal's radii. But it turned out in the study on collagen, that this is not strictly true and the limiting distances are to be reformulated. Using the analysis on the crystal structure data on small molecules, two sets of limiting distances of closest approach were obtained, named as normal and outer limits. Any specific conformation of a pair of linked peptide units could be generated for a given pair of (ϕ, ψ) using analytical methods and the distances between the various pairs of non-bonded atoms could be examined using these limiting distances. The conformation was then designated as fully allowed or partially allowed or disallowed depending on whether (i) all the distances were greater than the normal limit or (ii) one or more distances were less than the outer limit (short contact) or (iii) some distances lie between the two limits.

Ramachandran Map for *L*-residues

The above procedure has been applied for the entire range of $-180°$ to $180°$ of ϕ and ψ at specific intervals. It is worthwhile to remember that this idea was conceived of and the laborious calculations had to be carried out at a time (1960-62) when computers did not make their appearance on the scientific scene and the only sophisticated tool available was the electric desk calculator. Nevertheless the entire calculation was carried out. The next question that arose was how to represent the results. Since this was a two-parameter problem the obvious solution was a sort of x-y plot. Cutting the long story short, the final result is depicted in *Figure* 4, which has now universally come to be known as Ramachandran map. Any student of chemistry/biochemistry

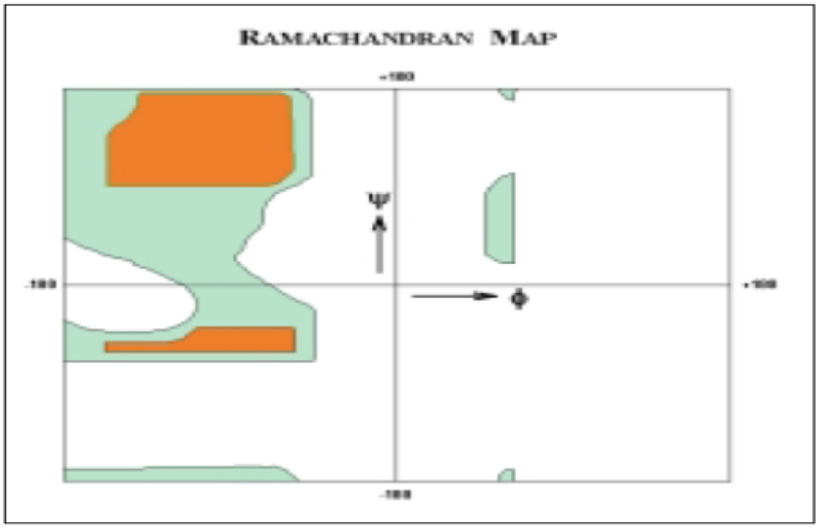

Figure 4. Ramachandran map for non-glycyl residues at the junction a-carbon atom. The fully allowed regions are shown in red shade, and the partially allowed regions in green shade.

will know that the amino acids can exhibit stereo-isomerism and the dominant isomer in proteins is *L*-amino acid. For the formulation of the map, the side chain *b*-carbon atom in *L*-configuration has been used.

Let us now take a brief survey of the Ramachandran map. The fully allowed regions are shown in red shade in *Figure* 4 and partially allowed regions in green shade. The rest are disallowed regions. There are two fully allowed regions, one on the top left and the other on the bottom left quadrants and two partially allowed regions, one on the left half and a small one on the right half.

Ramachandran Map and Protein Structures

When the Ramachandran map was first published (in 1963) not even a single protein structure had been solved. However, structures such as *a*-helix and *b*-sheet, proposed for fibrous proteins of kmef family and silk were available in literature in addition to a handful of small peptide crystal structures. It was very gratifying to note that the Rama-chandran angles for these, fall well within the allowed region. The map had to be kept in cold storage for a few more years before the crystal structures of proteins made their appearances on the scientific scene. At present, more than 15,000 structures are known (crystallographically solved, NMR and modelled) and with the enormous amount of computational power available, the Ramachan-dran angles can be computed and depic-ted on the backdrop of the Ramachandran map. One such example (chosen randomly) is given in *Figure* 5 for the protein penicillopepsin (acid hydrolase; PDB code 3APP), which has around 280 non-glycyl residues. The conformations can be seen to fall well within the allowed regions and this excellent agreement clearly brings out the strength of a systematic study based on a sound fundamental principle. At present, the map is used effectively as checkpoint during and after the crystal structure is solved.

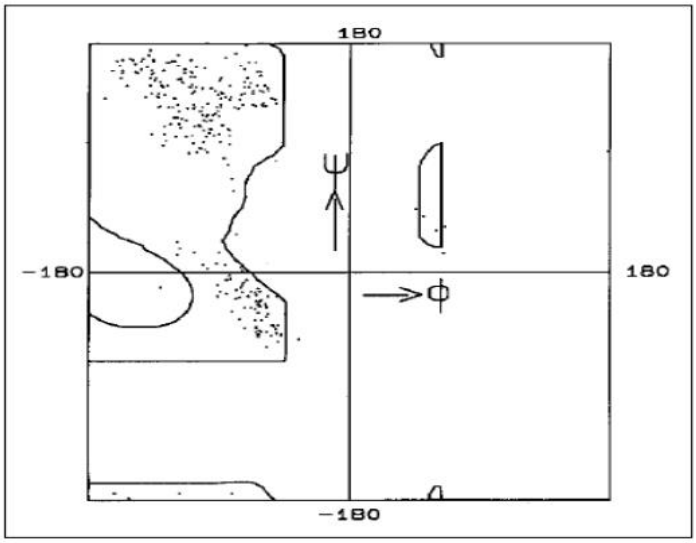

Figure 5. Ramachandran plot of amino acid residues in the protein, penicillopepsin (acid hydrolase; PDB code-3APP). Only the non-glycyl residues are plotted. The excellent agreement can be seen in that the points fall well within the outer limit allowed regions.

Utility value of the Ramachandran Map

Ramachandran map comes in as a handy tool in any analysis involving protein structure. The two important secondary structures that are found in any protein are (i) *a*-helices and (ii) *b*-strands (also known as extended strands) forming *b*-sheets.

Though, these are nowadays depicted through wonderful graphic representations (cartoon diagrams as they are called), conformational information can be visualised with the help of the Ramachandran map. *Figure* 6a and b show both the ribbon diagram (drawn using the software package RASMOL) and the Ramachandran plots for these two secondary structures. It is easily seen and of course true that the Ramachandran angles in a helical segment/extended strand are all clustered, but at different quadrants in the map. There are many such instances where the map can be effectively used in data analysis on proteins such as motif recognition, closeness between two structures, etc.

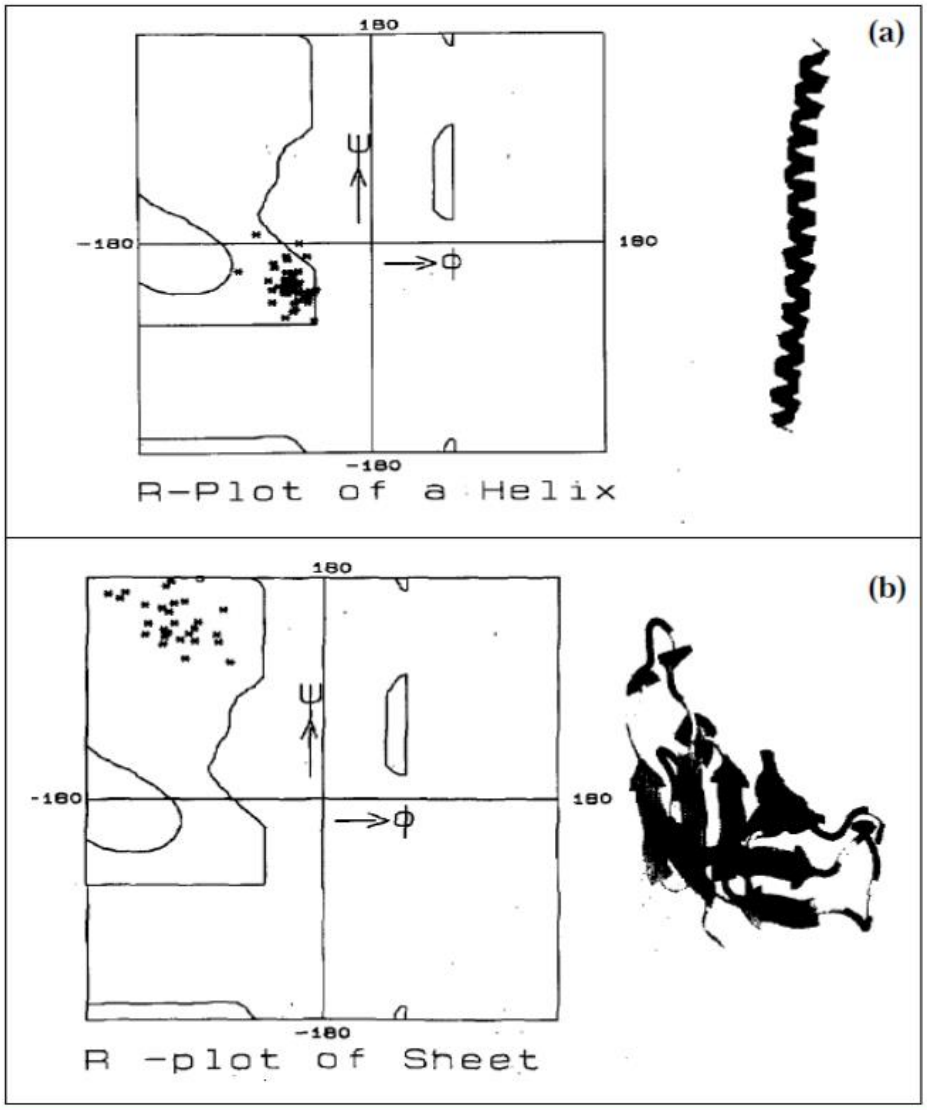

Figure 6. Clustering of points for regular secondary structures is a feature that can be easily brought out from Ramachandran plot. (a) Clustering in **a***-helical segment (around (–60°, –30°) of (*ϕ*,*ψ*), shown on the left, and ribbon diagram of the segment shown on the right) (b) Clustering in* **b***-sheet (shown on the left); (the ideal value for a fully extended chain is (–180°, 180°) of (*ϕ*,*ψ*) and the corresponding ribbon diagram is shown on the right). Note that the locations of the clusters for the secondary structures are located in different quadrants of the map.*

Apart from these two well-known secondary structures, there is a third one, which in many instances is responsible in bringing around the folded nature in proteins. These are known as *b*-turns and involves a system of three linked peptide units. Literature is full of many studies on these. Ramachandran and his group were the first ones to recognize the importance of such a motif as a part of a systematic study on hydrogen bonds in peptide segments.

Other Examples of Ramachandran Maps

Though the Ramachandran map was produced using the *L*-

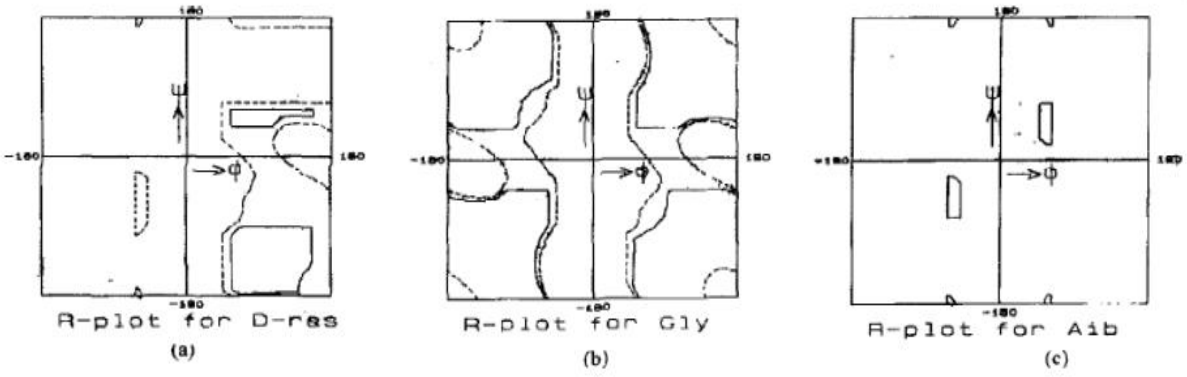

Figure 7. Ramachandran maps corresponding to (a) D-residue (b) glycyl residue and (c) Aib residue at the junction a-carbon atom ; The inverted appearance of the D-map as compared to the conventional Ramachandran map as well as the symmetrical nature of the allowed regions in (b) and (c) which arises due to the nonchiral nature of the side groups are observable.

configuration for the side chain *b*-carbon atom, it can easily be adopted for the *D*-residue. Ramachandran maps for *D*-, Gly and Aib (Alpha-amino-isobutyric acid – an unusual amino acid with two *b*-carbon atoms) residues are shown in *Figure* 7a-c, respectively. The huge conformational freedom enjoyed by the glycyl residues due to the absence of *b*-carbon atom and the high restriction imposed due to the presence of two *b*-carbon atoms has been brought out clearly in the Ramachandran maps. Similarly a plot of prolyl residues (not shown here) will show the severe geometric restriction imposed on one of the Ramachandran angle *f* (around –60° for *L*-Pro).

In addition to the map, Ramachandran made a number of isolated but highly significant contributions in the field of protein conformation. Mention can be made of the studies on prolyl residues, relative merits of different forms of potential functions for non-bonded interaction, hydrogen bonding potential function, helix with alternating *L*- and *D*-residues and many more. The concept of Ramachandran map first developed for proteins were also applied to other biopolymers such as nucleic acids and polysaccharides.