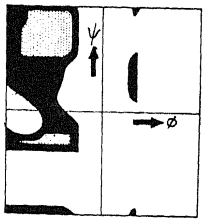

Special issue:
The Ramachandran map and
protein and peptide structures

# Glycyl residues in proteins and peptides: An analysis

**C. Ramakrishnan* and N. Srinivasan**
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

**Glycyl residue, simplest of all the residues, is well known for its conformational freedom. An analysis of the conformational and structural aspects of this residue occurring in proteins and peptides has been made making use of the Ramachandran ($\phi, \psi$) angles and their distribution. The chief observations are: (i) By and large, there is no bias for an amino-acid residue to precede or succeed Gly. (ii) The conformational points show clustering in the 'bridge' region. (iii) While in general, glycyl residue plays a passive role when it occurs in a helix and helps in helix propagation, it also acts as a helix breaker in some instances. (iv) It is a poor former of extended strands. (v) The conformational freedom is effectively used by Gly to prefer those positions in turns that are less favourable for non-glycyl residues. (vi) Analysis of doublet data reinforces the propagative tendency of glycyl residues. X-Gly doublet is a better turn former than Gly-Y doublet, (vii) Only one third of the glycyl residues are situated on the surface of the proteins. The results can be useful in modelling studies on proteins.**

ANALYSIS and documentation of protein backbone conformation are being widely carried out using the ($\phi, \psi$) angles. The 'time-tested' Ramachandran map[1] provides an excellent tool to study the general stereochemical feasibility of the peptide backbone. The secondary structural features such as $\alpha$-helix, $\beta$-sheet and $\beta$-turn can be represented by one or two points in the ($\phi, \psi$) space[2] and this aspect enables one to quickly recognize such structural motifs in proteins. Thus the regularities in three-dimensional complexity of globular proteins are reduced on to a comprehendable two-dimensional Ramachandran ($\phi, \psi$) plane. In the literature, the Ramachandran plot is commonly used to assess the stereochemical quality of any modelled protein. Glycyl residues (Gly), in most cases are not considered separately in spite of its high degree of conformational freedom.

Over the years, extensive attempts have been made to understand the significance of the role of Gly in various synthetic polypeptides and fibrous proteins like collagen and silk[3]. The poly(Gly) I adopts a $\beta$-sheet structure[4] and poly(Gly) II a three-fold left-handed triple-helical structure[5]. Silk (*Bombyx mori*) contains nearly 50% of Gly and the presence of glycine in almost every alternate position in the sequence facilitates close packing of $\beta$-sheets[6]. Gly occupies every third location of the collagen chains. The chains of the triple-helical collagen get closely packed due to the glycyl residues[7].

Glycyl has some nearly direct roles in the function of some peptides and proteins. For example, the sequence Arg–Gly–Asp–Ser is a part of the cell attachment domain of fibronectin[8]. It is also involved in the debatable mechanism of binding of carboxypeptidase with its ligand[9-11] and in the activation of oncogenic p21 protein where Gly-12 is crucial for activity[12].

In view of these interesting sequence and structural characteristics of Gly in various systems, the present study is motivated towards extracting its conformational peculiarities and positional preferences in crystal structures of proteins and peptides. The Ramachandran ($\phi, \psi$) angles are extensively used as a potential tool to investigate the conformational aspects.

## Glycyl residues in the primary structure

The occurrence of certain sequence patterns involving glycyl residues is well known for fibrous proteins such as collagen and silk. In the former, not only does glycyl form about one third of the total number of residues but occur as every third residue[7]. So also in $\beta$-proteins glycyl occurs more as . . . Gly-X–Gly-X–Gly-X. . . .

In order to find out whether there are any amino acids which have extreme bias of occurrence adjacent to glycyl residues in globular proteins, a data set comprising of 119 non-homologous proteins chosen from the protein data available in the Brookhaven Protein Data Bank[13] has been used. This set contains 2135 Gly occurring in these proteins. Considering a

*For correspondence.