Protein Engineering vol.12 no.7 pp.523–526, 1999

SHORT COMMUNICATION

Graphical representation of the salient conformational features of protein residues

Debnath Pal and Pinak Chakrabarti[1]

Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India

[1]To whom correspondence should be addressed

A composite plot for depicting in two dimensions the conformation and the secondary structural features of protein residues has been developed. Instead of presenting the exact values of the main- and side-chain torsion angles (ϕ, ψ and χ_1), it indicates the region in the three-dimensional conformational space to which a residue belongs. Other structural aspects, like the presence of a *cis* peptide bond and disulfide linkages, are also displayed. The plot may be used to recognize patterns in the backbone and side-chain conformation along a polypeptide chain and to compare protein structures derived from X-ray crystallography, NMR spectroscopy or molecular modelling studies and also to highlight the effect of mutation on structure.

Keywords: conformation/protein engineering/protein structural comparison/secondary structure/structural plot

Introduction

Visualization of protein structures is an important aspect in any study pertaining to structure–activity correlation. A three-dimensional atomic model in stereo provides the most precise representation, but owing to the large number of atoms involved, the protein fold is usually represented as a 'cartoon' or ribbon drawing (Richardson, 1981). Two-dimensional graphs can also be used to portray three-dimensional information in an easily comprehensible way. After the elucidation of the Ramachandran map (Ramachandran *et al.*, 1963) relating the torsion angles ϕ and ψ, the variation of the structural features along the main-chain, especially in turns, has been depicted by joining consecutive residues by arrows on the plot (Venkatachalam, 1968). However, such drawings are often confusing when a large number of points are present on the map. To avoid such confusion, chain plots were introduced (Srinivasan *et al.*, 1975) in which one or more conformational parameters were plotted versus the sequence. Balasubramanian (1977) plotted the ϕ and ψ values together on the *y*-axis (and connected by a vertical line) against residue numbers along the *x*-axis, so that the resultant plot is an unlinked series of parallel line segments. A variant of this is the linked $\phi\psi$ chain plot (McClain and Erickson, 1995) in which instead of plotting both the ϕ and ψ angles on the same vertical line, they are shifted horizontally and then connected to each other and to points due to neighboring residues by slanting line segments. This gives a contiguous picture of protein backbone conformation, and such a trace is more amenable for structural comparison. However, all these plots are impervious to any information on side-chain conformation (χ_1). Even if χ_1 values are included (Srinivasan and Yathindra, 1978; Sheriff *et al*., 1987), the plots lack clarity.

We have recently shown (Chakrabarti and Pal, 1998) the interdependence of the main- and side-chain torsion angles (ϕ, ψ and χ_1), and there is also a need for better understanding of the role of the side-chain in the packing and stability of the protein fold. As such, an unambiguous representation of χ_1 is as important as the $\phi\psi$ chain trajectory, and a plot discussed here meets this requirement.

Materials and methods

The structural coordinates from the Brookhaven Protein Data Bank, PDB (Bernstein *et al*., 1977) were used to obtain the conformational parameters using the PROCHECK suite of programs (Laskowski *et al.*, 1993). The 'secstr' module of the latter was modified to label the disulfide linkages and only relevant parameters were written to the output '.rin' file. The χ_1 angle of Val was also modified to conform with the convention used for Ile and Thr (Chakrabarti and Pal, 1998). A computer program CONFPLOT was written in FORTRAN 77 to generate a postscript file for the plot.

The *y*-axis of the plot is divided into four major bands corresponding to the four regions of the Ramachandran map containing non-overlapping clusters of ϕ, ψ points (Chakrabarti and Pal, 1998) (Table I). Each major region is further subdivided into four groups based on χ_1 (Table II).

Each residue, depending on its ϕ, ψ, χ_1 angles, has a distinct position (open circle) on the plot and the consecutive points are connected. The terminal residues, with either ϕ or ψ missing, are shown as filled circles on the base line (for the first residue) or at the same level as the penultimate residue (for the last residue). If there is a break in the backbone, the residue after the break is shown at the same level as the previous residue, but not connected to it. If, due to missing atoms, χ_1 could not be calculated, the residue is shown at the bottom of the band corresponding to the main chain. There are a number of panels below the *x*-axis. The first indicates the sequence number of the residues as given in the PDB file. The second row indicates the amino acid type (non-natural amino acid residues are indicated as U), while the third shows the secondary structural notifier as given in the output of PROCHECK. To indicate a *cis* peptide bond, a filled square is drawn at the center of the line joining the two residues. Cysteine residues involved in disulfide bridges are labeled a, b, . . ., corresponding to the serial number of the bridge, and these are shown on the top band. Metal-binding residues can be indicated in the same fashion.

Multi-chain files can easily be processed as one specifies the range to be plotted by providing the serial numbers of the residue records in the '.rin' file. Any sequence length can be handled by the program. However, for a long chain, if the labels along the *x*-axis become cluttered, one can split it into smaller segments to be plotted separately (a maximum of nine allowed on a single page). If one is interested in the trace only, it is also possible to put labels at an interval on the *x*-axis.

For overlaying the second (and subsequent) plots (drawn

Table I. Designations for different regions in the Ramachandran plot

Region	φ (°)	ψ (°)
A	–180 to 0	–120 to 60
B	–180 to 0	60 to 240
L	0 to 180	–90 to 90
R	0 to 180	90 to 270

Table II. Plot labels and the corresponding side-chain conformations

Sub-region	χ_1 (°)
x	for Gly and Ala with no χ_1
t	120 to 240
g^+	–120 to 0
g^-	0 to 120

with broken lines with varying gap widths) the required number of '.rin' files are read. Polypeptide chains with similar structures but different sequences (and residue numbers) can be plotted together, but the prior information on sequence alignment and gap positions have to be provided.

Results and discussion

Plot description

For clarity, instead of plotting the three torsional angles, we have identified each residue by the well-defined regions in the three-dimensional conformational space in which it occurs. So the φ, ψ values, depending on the location on the Ramachandran map, are replaced by a single label A, B, L or R, corresponding to four regions in the Ramachandran map—the first three are named after the prominent secondary structural element each encompasses (A, α-helix; B, β-strand; L, left-handed α-helix; R, the remaining region). Likewise, the side-chain conformation is shown not by the actual χ_1 value, but by three idealized states, *gauche* (g^+ and g^-) and *trans* (*t*); Gly and Ala, with no χ_1, are shown against '*x*' along the *y*-axis. Besides the conformational data, the secondary structural features of residues, disulfide linkages are also given. If needed, additional panels can be inserted to show residues involved in function, subunit association or substrate binding. A few representative plots are presented in Figures 1–3.

Main-chain conformation and secondary structure

The identification of a stretch of backbone with a right- or left-handed α-helical and β-strand conformation is very easy as these appear in separate bands. However, in going from one residue to the next, if the change in conformation does not involve a shift from one to another in the four distinct regions of the Ramachandran map, it does not show up in the plot. As a result, it cannot distinguish between α- and 3_{10}-helices or parallel and antiparallel β-sheets which can be identified only if the φ, ψ values are plotted, but even then, as the differences are small the patterns in the plot would be nearly identical (McClain and Erickson, 1995). Of the two widely occurring β-turns, the type II turn with the two central residues having φ, ψ combinations of [(–60,120) and (80,0)] involving a B→L change is easily recognizable, but not the type I turn [(–60,–30) and (–90,0), A→A change]. Such ambiguities can be resolved from the secondary structural information provided below the residue names. The plot can very elegantly display a change in conformation from one

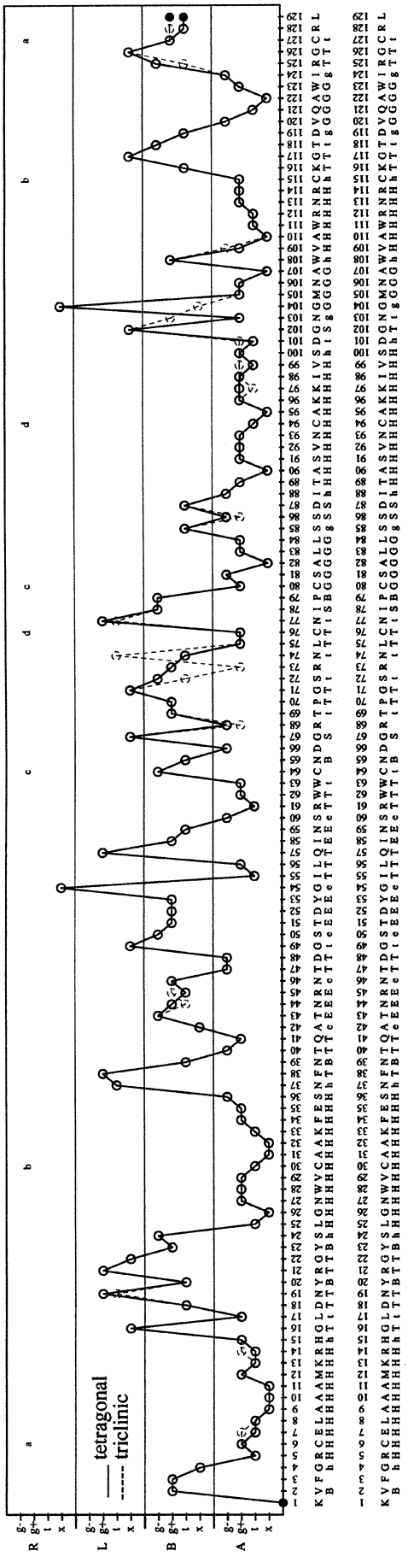

Fig. 1. CONFPLOT for the structures of hen egg white lysozyme in the tetragonal and triclinic crystal forms [PDB files 193L (Vaney,M.C., Maignan,S., Ries-Kautt,M. and Ducruix,A., manuscript in preparation) and 2LZT (Ramanadham *et al.*, 1990)]. a, b, c and d are the designations of the four disulfide linkages.

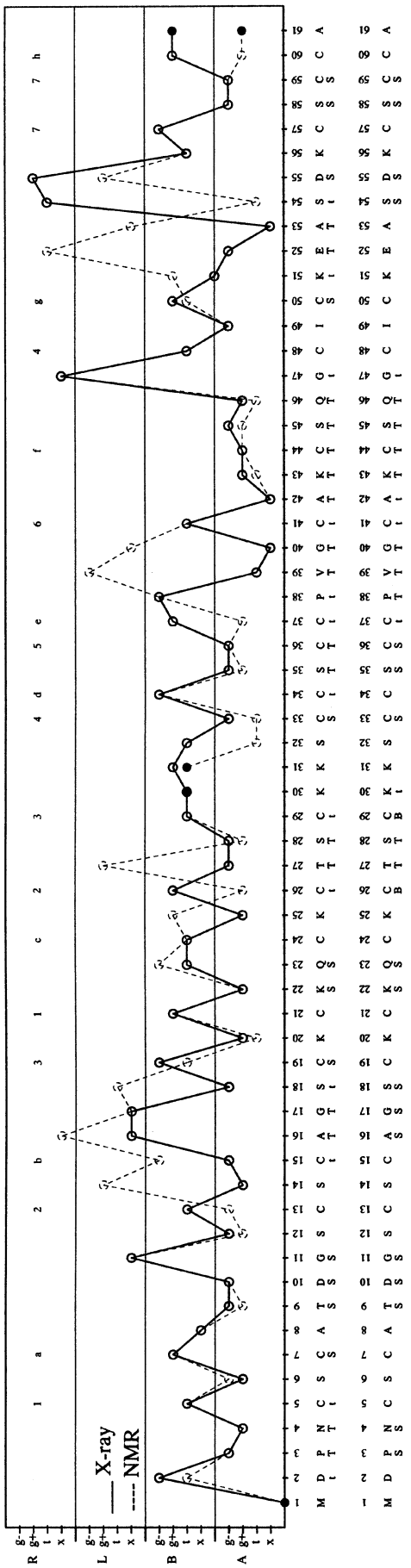

Fig. 2. Comparison of the X-ray and NMR structures of Cd,Zn-metallothionein [PDB files 4MT2, 1MRT and 2MRT (Braun *et al.*, 1992)]. Metal binding ligands are identified as 1, 2, . . ., 7 depending on the serial number of the cations [the first three constitute the $CdZn_2(S\gamma)_9$ cluster, and the remaining four, the $Cd_4(S\gamma)_{11}$ cluster] they are coordinated to, or as a (coordinating simultaneously to 1 and 2), b (2,3), c (1,3), d (4,5), e (5,6), f (4,6), g (5,7) and h (6,7). The NMR structure has a break in the peptide bond between residues 30 and 31.

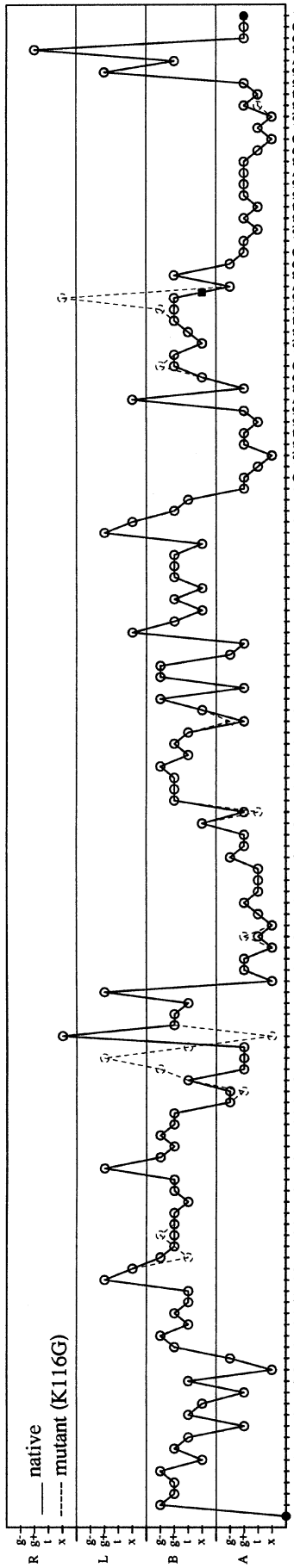

Fig. 3. Effect of the Lys116→Gly mutation on the native structure of staphylococcal nuclease [PDB files 1STN (Hynes and Fox, 1991) and 1KAB (Hodel *et al.*, 1993)]. The native structure has a *cis* peptide bond between residues Lys116 and Pro117, which is observed as the *trans* isomer in the mutant.

region of the Ramachandran map to another, especially when the sequence of residues do not constitute any regular secondary structure. Likewise, the occurrence in region R, which is not very common for a non-Gly residue, is clearly shown, and the context in the primary and secondary structures in which such a residue is located can be analysed from the plot.

Side-chain conformation

One of the motivations for designing the plot was to see if there is any pattern in the side-chain conformation as one moves along the polypeptide chain—a matter which has not received adequate attention in the literature. For example, considering the four distinct possibilities (no χ_1 and the three conformational states of χ_1) in a given region, one may ask the question if a residue with no χ_1 occurs at any specific position, or for other residues, if any combination of the conformational states is preferred in adjacent positions or with a fixed interval.

Structural comparison

The plot provides an easy visual comparison of protein structures, as shown in Figure 1, for the two crystal forms of hen egg white lysozyme. There are changes not only in the side-chain conformation of a few residues, but also in two stretches of the main chain (residues 73–74 and 103–104). The changes in the side-chain conformation usually get overlooked as most of the publications report the root mean square deviation calculated using backbone atoms only. However, we have recently shown that the residues having multiple side-chain conformation usually have their main-chain torsion angles restricted (Chakrabarti and Pal, 1998). In another example, Figure 2 compares the structures of metallothionein (Braun *et al.*, 1992) obtained using X-ray and NMR methods, and indicates that the non-ligand residues have significant conformational differences.

We have been interested in the structure and dynamics around the *cis* peptide linkages (Pal and Chakrabarti, unpublished work) and came across a study on a Lys116→Gly mutation in staphylococcal nuclease (Hodel *et al.*, 1993). The peptide bond between residues Lys116 and Pro117 is *cis* in the wild-type structure but becomes *trans* in the mutant. The structural changes were reported to be limited not only in the region of mutation (111–119) but also in regions further away (44–51). The consideration of side-chain conformation in the plot indicates that there are many more alterations in the structure. The plot can thus highlight changes accompanying protein engineering experiments.

Since the plot is easy to visualize, additional parameters can be displayed for other applications. One can use this plot to highlight the disordered regions of a molecule to find out if there is any link with the structure and the sequence they are embedded in. NMR models which give an ensemble of structures can be plotted to visualize variation along the sequence. Flexibility associated with the surface accessibility may also be assessed from these plots.

Conclusion

The Ramachandran map is one of the most widely used tools for protein structure analysis. Our plot relates directly to the map and provides further information on the side-chain conformation. Since the nature of the side-chain is the differentiating factor among the amino acid residues, an analysis of protein structure vis-à-vis the side-chain will be facilitated by the highlighted display of its conformation in the plot. Additional structural information, as to which residues are involved in disulfide bridges, metal binding and secondary structural features (and others that a user may want to include), should make the plot useful as a molecular 'finger print.'

The software is available for academic use from the authors, and the file (pub/pinak/confplot/confplot.tar.gz) can be downloaded from boseinst.ernet.in.

Acknowledgements

Financial support was provided by the Department of Science and Technology, Department of Biotechnology and the Council of Scientific and Industrial Research.

References

Balasubramanian,R. (1977) *Nature*, **266**, 856–857.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.,Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
Braun,W., Vasak,M., Robbins,A.H., Stout,C.D., Wagner,G., Kägi,J.H.R. and Wüthrich,K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10124–10128.
Chakrabarti,P. and Pal,D. (1998) *Protein Engng*, **11**, 631–647.
Hodel,A., Kautz,R.A., Jacobs,M.D. and Fox,R.O. (1993) *Protein Sci.*, **2**, 838–850.
Hynes,T.R. and Fox,R.O. (1991) *Proteins Struct. Funct. Genet.*, **10**, 92–105.
Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.
McClain,R.D. and Erickson,B.W. (1995) *Int. J. Pept. Prot. Res.*, **45**, 272–281.
Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
Ramanadham,M., Seiker,L.C. and Jensen,L.H. (1990) *Acta. Crystallogr.*, **B46**, 63–69.
Richardson,J.S. (1981) *Adv. Prot. Chem.*, **34**, 167–339.
Sheriff,S., Hendrickson,W.A. and Smith,J.L. (1987) *J. Mol. Biol.*, **197**, 273–296.
Srinivasan,R., Balasubramanian,R. and Rajan,S.S. (1975) *J. Mol. Biol.*, **98**, 739–747.
Srinivasan,A.R. and Yathindra,N. (1978) *Biopolymers*, **17**, 1595–1600.
Venkatachalam,C.M. (1968) *Biopolymers*, **6**, 1425–1436.

Received February 19, 1999; revised April 14, 1999; accepted April 14, 1999