

# Termination of right handed helices in proteins by residues in left handed helical conformations

H.A. Nagarajaram, R. Sowdhamini, C. Ramakrishnan and P. Balaram

*Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India*

Received 27 January 1993

An analysis of 636 helical segments, ranging in length from 4 to 32 residues, from 123 independent protein crystal structures reveals that helix termination by residues in left handed ( $\alpha_L$ ) helical conformations is a common occurrence. Gly and Asn residues are the most frequent  $\alpha_L$  helix terminators, with the former having a very high propensity to adopt such conformations. The  $\alpha_R$ - $\alpha_R$ - $\alpha_R$ - $\alpha_L$  segment at the C termini of protein helices often possesses a  $6 \rightarrow 1$  ( $\pi$ -type) hydrogen bond between the CO of residue  $i$  and the NH of residue  $i + 5$  with residue  $i + 4$  occurring in the  $\alpha_L$  conformation. A stereochemical analysis of 216 examples shows that in 62 cases the  $6 \rightarrow 1$  hydrogen bond is absent. The present analysis provides a quantitative measure of the propensity of the 20 amino acids to adopt  $\alpha_L$  helix terminating conformations.

Helix termination;  $6 \rightarrow 1$  Hydrogen bonds; Protein conformation; Protein data analysis

## 1. INTRODUCTION

$6 \rightarrow 1$  ( $C_{16}$  or  $\pi$ -type) hydrogen bonds involving the carbonyl oxygen of residue  $i$  and the amide hydrogen of residue  $i + 5$  occur infrequently in proteins, as compared to the more widely observed  $4 \rightarrow 1$  ( $C_{10}$ ) and  $5 \rightarrow 1$  ( $C_{13}$ ) hydrogen bonds which are found in  $\beta$ -turns/ $3_{10}$ -helices and  $\alpha$ -helices, respectively [1]. Helices in proteins often terminate with a reversal of helix sense at the C terminus, the helix breaking residue adopting a left-handed,  $\alpha_L$  ( $\phi \approx +50^\circ$ ,  $\psi \approx +50^\circ$ ) conformation. This feature generally results in the formation of a  $6 \rightarrow 1$  hydrogen bond, along with a  $5 \rightarrow 2$  ( $C_{10}$ ) hydrogen bond, with residue 5 occurring in the  $\alpha_L$  conformation, as first reported by Schellman [2]. A subsequent analysis, using a data set of 40 protein structures, revealed the occurrence of  $6 \rightarrow 1$  hydrogen bonded features, with concomitant  $5 \rightarrow 2$  hydrogen bond formation in isolated hairpins, referred to as 'paper-clips' [3]. Residue 5 adopts the  $\alpha_L$  conformation and is predominantly Gly and to a lesser extent Asn [3], an observation also highlighted in earlier studies [2]. A recent observation of the  $6 \rightarrow 1/5 \rightarrow 2$  hydrogen bonded  $\pi$ -turn at the C terminus of the crystalline, helical peptides, with the achiral  $\alpha$ -aminoisobutyryl (Aib) residue adopting the  $\alpha_L$  conformation [4], stimulated a detailed re-examination of this helix terminating structural feature in proteins. This communication presents the results of an analysis based

on a large data set of 636 helices, of length  $\geq 4$  residues, observed in 116 independent protein crystal structures. The overwhelming majority of helix terminating residues with  $\alpha_L$  conformations in proteins are Gly residues. Asn residues also show an appreciable propensity for this conformation. This analysis provides a quantitative estimate of the propensity of the 20 amino acid residues to occur in this helix terminating structural feature.

## 2. MATERIALS AND METHODS

A data set of 123 protein structures determined at resolution  $\leq 2$  Å were extracted from the Brookhaven Protein Data Bank [5]. The data set consists of largely non-homologous structures and will be detailed elsewhere. Helices were identified using the criterion that four successive backbone  $\phi$ ,  $\psi$  values [6] should lie within the region defined by the limits ( $-120^\circ \leq \phi \leq 0^\circ$ ;  $-60^\circ \leq \psi \leq 0^\circ$ ; a rectangular box in the ( $\phi$ - $\psi$ ) map). The helix data set so formed consisted of 636 helical segments ranging in length from 4 to 32 residues. The residue immediately succeeding the helix at the C-terminal end was designated as the 'terminator' residue and  $\phi$ ,  $\psi$  values were computed. If these  $\phi$ ,  $\psi$  values lie within the box as defined by the limits ( $0^\circ \leq \phi \leq 120^\circ$ ;  $-40^\circ \leq \psi \leq 120^\circ$ ), then the conformation of that corresponding residue was considered as left-handed helical ( $\alpha_L$ ) and such residues along with other 'terminator' residues were picked out for further study as outlined below.

Propensity ( $p_{\alpha_L}^i$ ) of a residue  $i$  to occur at the C-terminal end of the helices (terminator position) having  $\alpha_L$  conformation was computed as follows:

$$\text{Propensity, } p_{\alpha_L}^i = \frac{f_{\alpha_L}^i}{F_i^i}$$

where,  $f_{\alpha_L}^i$  is the % occurrence of a 'terminator' residue  $i$ , in  $\alpha_L$  conformation and  $F_i^i$  is the % occurrence of a residue  $i$  that is generally found

Correspondence address: P. Balaram, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. Fax: (91) (812) 341 683.

at the terminator position. These terms were in turn evaluated using the following,

$$f_{\alpha_L}^i = \frac{n_{\alpha_L}^i}{N_{\alpha_L}} \text{ and } F_i^t = \frac{n_i^t}{N_t}$$

here,  $n_{\alpha_L}^i$  = the number of occurrences of the residue  $i$  in  $\alpha_L$  conformation,  $n_i^t$  = the number of occurrences of the  $i^{\text{th}}$  residue at the helix termini in the entire data set of helices,  $N_{\alpha_L}$  = total number of residues having  $\alpha_L$  conformation and  $N_t$  = total number of residues found at the helix termini, in the data set.

### 3. RESULTS AND DISCUSSION

Fig. 1a shows the distribution of  $\phi$ ,  $\psi$  values at the 'terminator' residue. Of the 636 residues marked, 208

are Gly residues. Fig. 1b and 1c provide the observed  $\phi$ ,  $\psi$  distributions for the individual residues Gly and Asn. The clustering of Gly residues in  $\alpha_L$  conformations at helix C-termini is dramatic and may be compared with overall distributions for Gly residues in proteins, reported earlier [7]. The clustering in the  $\alpha_L$  region is less pronounced for Asn. Fig. 1d shows the propensity of each of the 20 amino acids to occur in the  $\alpha_L$  conformation at the helix termini. Clearly the propensities of the non-Gly/non-Asn residues to occur in  $\alpha_L$  conformations at the C-termini of helices is very low with several residues being practically unrepresented. Proline, of course, is stereochemically prohibited from adopting  $\alpha_L$  conformation since  $\phi$  for L-Pro is restricted to  $-60^\circ \pm 20^\circ$ . The

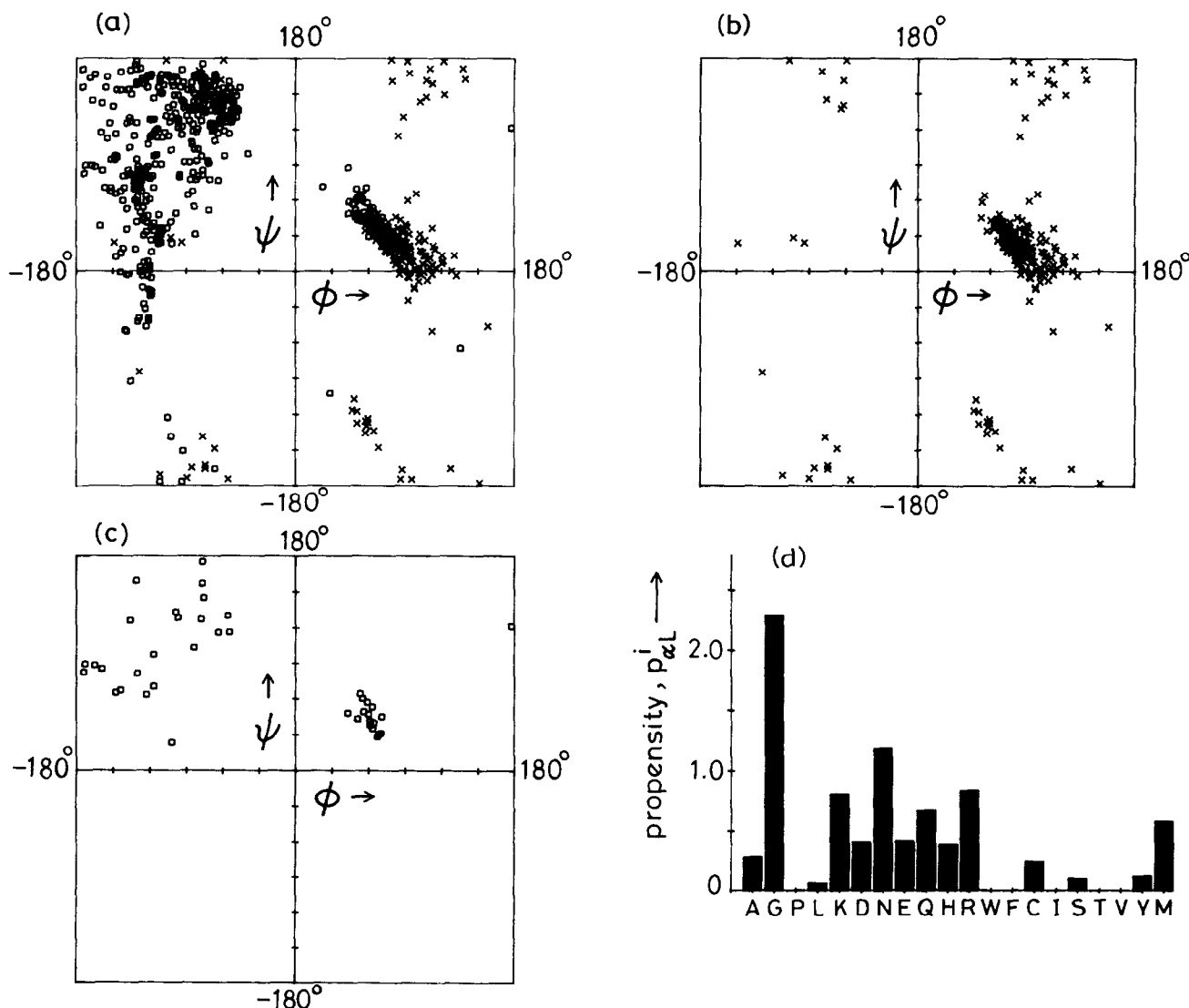


Fig. 1. (a) The  $\phi$ ,  $\psi$  distribution of 'terminator' residues that are found at the C-terminal end of the 636 helices. The non-glycyl residues are marked with 'o', while glycyl residues are marked with 'x'. (b) The  $\phi$ ,  $\psi$  distribution in the case of terminator glycyl residues. In the majority of the examples of Gly, being the terminator residue, the conformation of Gly is  $\alpha_L$  (78%). (c) The  $\phi$ ,  $\psi$  distribution in the case of terminator Asn residues. Relatively equal distribution of points can be seen in the  $\alpha_L$  region as compared to the rest of the map. There are 16 Asn examples which have  $\alpha_L$  conformations, while 24 Asn residues occur elsewhere in the ( $\phi$ - $\psi$ ) map. (d) The propensity value of each of the 20 amino acids, to occur at the terminator position, having an  $\alpha_L$  conformation. Notable peaks that are  $> 1.0$  are due to Gly (2.29) and Asn (1.18). Residues which are totally absent are Pro, Trp, Phe, Ile, Thr and Val.

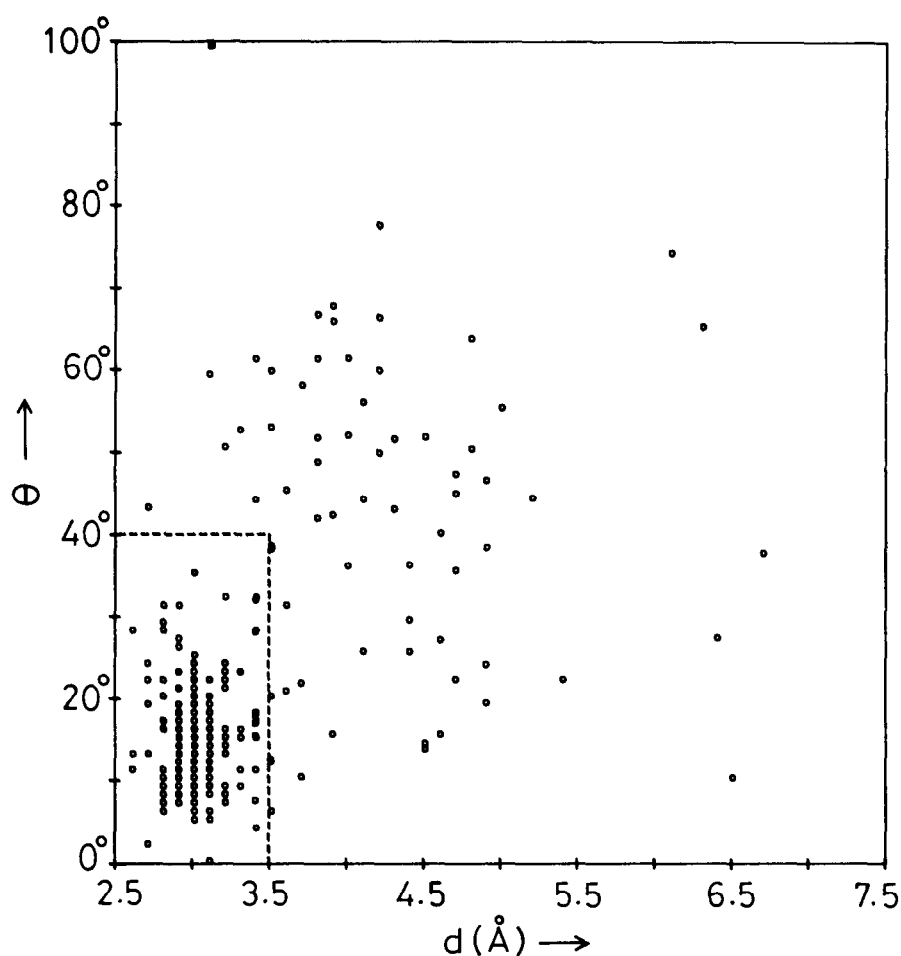


Fig. 2. The plot of ( $d$ - $\theta$ ) parameters of the  $6 \rightarrow 1$  hydrogen bond ( $\pi$ -type), where  $d$  is the distance between  $N_{i+5}$  to  $O_i$  and  $\theta$  is the angle  $H_{i+5} - N_{i+5} \dots O_i$ . A rectangular box is drawn enclosing the region spanned by  $2.5 \leq d \leq 3.5$  Å and  $0^\circ \leq \theta \leq 40^\circ$ . Values which lie within this box have been considered as hydrogen bonded (154 examples). The number of examples that fall outside the box is 62.

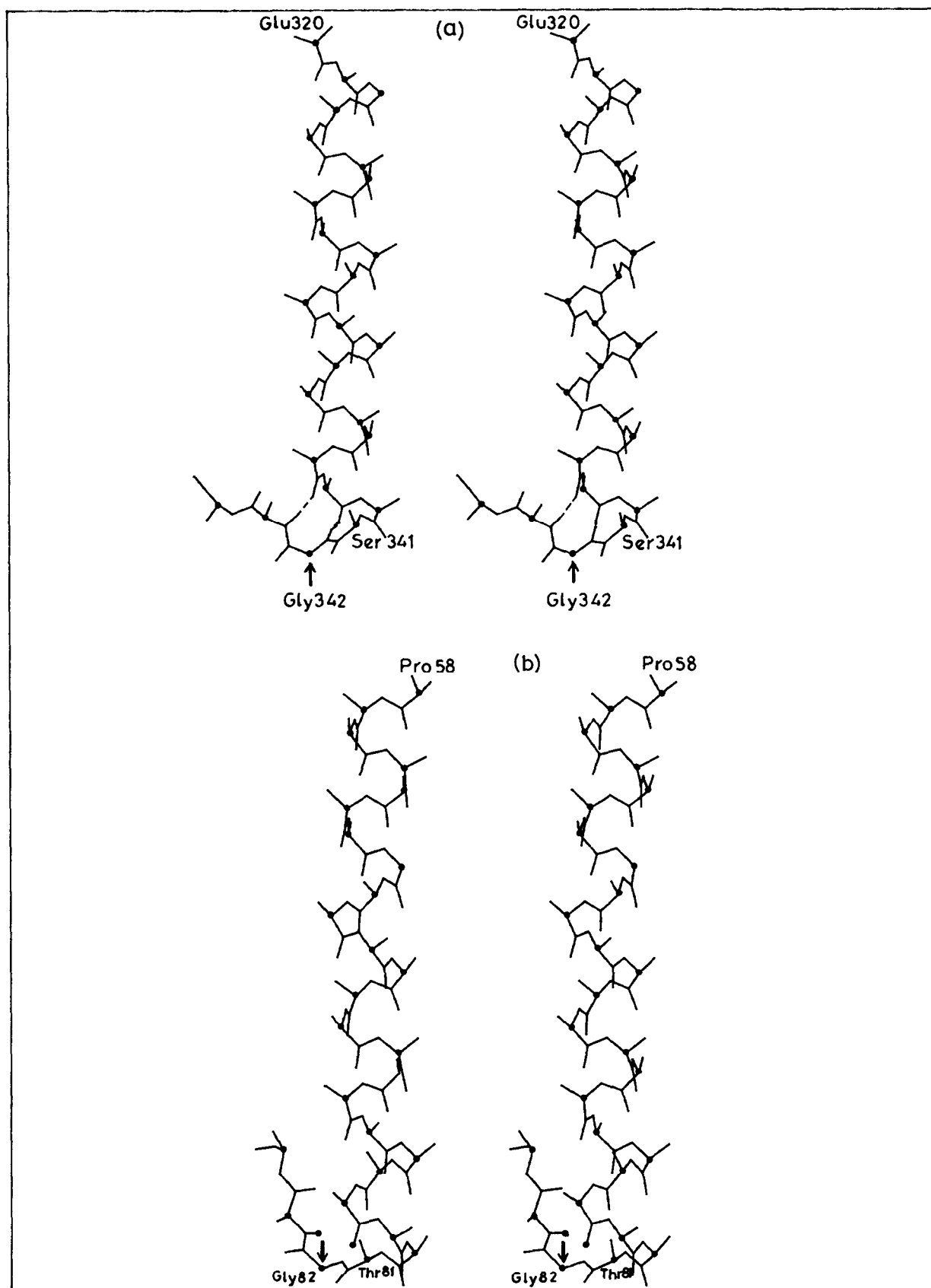
Table I  
Distribution of  $6 \rightarrow 1$  ( $\pi$ ) hydrogen bonds at helix termini involving  $\alpha_L$  terminators<sup>a</sup>

Residue ( $p_{\alpha_L}$ ) <sup>b</sup>	Number of occurrences <sup>c</sup>		Residue ( $p_{\alpha_L}$ ) <sup>b</sup>	Number of occurrences	
	With $6 \rightarrow 1$ hydrogen bond	No $6 \rightarrow 1$ hydrogen bond		With $6 \rightarrow 1$ hydrogen bond	No $6 \rightarrow 1$ hydrogen bond
Ala (0.28)	0	3	Arg (0.84)	5	1
Gly (2.29)	118	44	Trp (0.00)	0	0
Pro (0.00)	0	0	Phe (0.00)	0	0
Leu (0.06)	0	1	Cys (0.24)	1	0
Lys (0.81)	6	2	Ile (0.00)	0	0
Asp (0.41)	6	0	Ser (0.10)	0	1
Asn (1.18)	12	4	Thr (0.00)	0	0
Glu (0.42)	0	1	Val (0.00)	0	0
Gln (0.67)	3	2	Tyr (0.12)	0	1
His (0.39)	2	2	Met (0.58)	1	0

<sup>a</sup> Number of helices from 116 independent protein crystal structures.

<sup>b</sup> Values in parentheses correspond to the propensity of a residue to adopt  $\alpha_L$  conformations at helix termini. Values  $> 1.0$  indicate a preference for  $\alpha_L$  helix terminator conformations.

<sup>c</sup> The number of occurrences refers to the number of  $\alpha_R$ - $\alpha_R$ - $\alpha_R$ - $\alpha_L$  conformations at helix termini.



←

Fig. 3. Stereo diagrams of the two representative examples of helices that terminate with a residue in  $\alpha_L$  conformation (the side chains are shown only up to the  $C^\beta$  atom). (a) The helix-spanning residues Glu-320 to Ser-341 in glycolate oxidase (PDB code 1GOX). Gly-342 ( $\phi = +84^\circ$ ,  $\psi = +41^\circ$ ) is the terminator (indicated by an arrow). There are two hydrogen bonds, formed between the residues Met-338 to Cys-343 ( $\pi$ -type) and Ala-339 to Gly-342 ( $\beta$ -type) (shown by broken lines). (b) The helix spanning residues Pro-58 to Thr-81 in leghemoglobin (PDB code 2LH7). Gly-82 ( $\phi = +99^\circ$ ,  $\psi = -11^\circ$ ) is the  $\alpha_L$  terminator (indicated by an arrow). In contrast to the helix shown in Fig. 3a, there is no  $\pi$ -type hydrogen bond.

$\beta$ -branched residues Val, Ile and Thr are totally missing from the list of  $\alpha_L$  terminators, as are the aromatic residues Phe and Trp.

A further examination of the C terminus local structural feature in helices,  $\alpha_R$ - $\alpha_R$ - $\alpha_R$ - $\alpha_L$  conformations for the  $i + 1$  to  $i + 4$  segment, revealed that a large fraction of the observed examples did possess a  $6 \rightarrow 1$  hydrogen bond between the  $(CO)_i$  and  $(NH)_{i+5}$  groups (Table I). Fig. 2 shows the distribution of the structural parameters viz.,  $d$  (the distance from  $N_{i+5}$  to  $O_i$ ) and  $\theta$  (the angle  $H_{i+5}$ - $N_{i+5}$ - $O_i$ ) used to evaluate hydrogen bond geometry for the  $6 \rightarrow 1$  (i.e.,  $i + 5 \rightarrow i$ )  $\pi$ -type hydrogen bond. Points lying outside the rectangular box correspond to cases where a good hydrogen bond is not formed. Of the 162 examples involving Gly at position  $i + 4$  ( $\alpha_L$ ), 118 can be classified as  $6 \rightarrow 1$  hydrogen bonded. In the case of Asn, 12 out of the 16 cases possessed  $6 \rightarrow 1$  hydrogen bonds. Fig. 3 provides illustrations of both types of helix terminating segments. Of the 636 helices examined there are 216 examples where the terminating residue is in an  $\alpha_L$  conformation. In 75% of these cases the terminating residue is Gly and in 7.4% the residue is Asn. This suggests that C-terminal helix termination signals are well defined and algorithms for their identification might be usefully incorporated into protein secondary structure prediction strategies, particularly for the identification of helix stop signals [8,9].

The incorporation of unusual residues with high propensities for  $\alpha_L$  conformations (cf. D-residues) at appropriate sites might facilitate predictable helix termination in de novo protein design approaches and in protein engineering [10].

*Acknowledgements:* Financial support from the department of Science and Technology, Government of India, is gratefully acknowledged.

## REFERENCES

- [1] Baker, E.N. and Hubbard, R.E. (1984) *Prog. Biophys. Molec. Biol.* 44, 97-179.
- [2] Schellman, C. (1980) in: *Protein Folding* (R. Jaenicke, ed.) Elsevier, Amsterdam, pp. 53-61.
- [3] Milner-White, E.J. (1988) *J. Mol. Biol.* 199, 503-511.
- [4] Flippen-Anderson, J.L., Karle, I.L., Uma, K. and Balam, P. (1993) *Int. J. Peptide Protein Res.* (submitted).
- [5] Bernstein, F.C., Koetzle, T.K., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Mitsuo, T. (1979) *J. Mol. Biol.* 112, 535-552.
- [6] Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. and Balam, P. (1991) in: *Molecular Conformation and Biological Interactions* (P. Balam and S. Ramaseshan, eds.) Indian Academy of Sciences, Bangalore, India, pp. 59-73.
- [7] Ramakrishnan, C. and Srinivasan, M. (1987) *Int. J. Peptide Protein Res.* 29, 629-637.
- [8] Richardson, J.S. and Richardson, D.C. (1988) *Science* 240, 1648-1652.
- [9] Presta, L.G. and Rose, G.D. (1988) *Science* 240, 1632-1641.
- [10] Balam, P. (1992) *Curr. Opin. Struct. Biol.* 2, 845-851.