

# De novo protein design: Crystallographic characterization of a synthetic peptide containing independent helical and hairpin domains

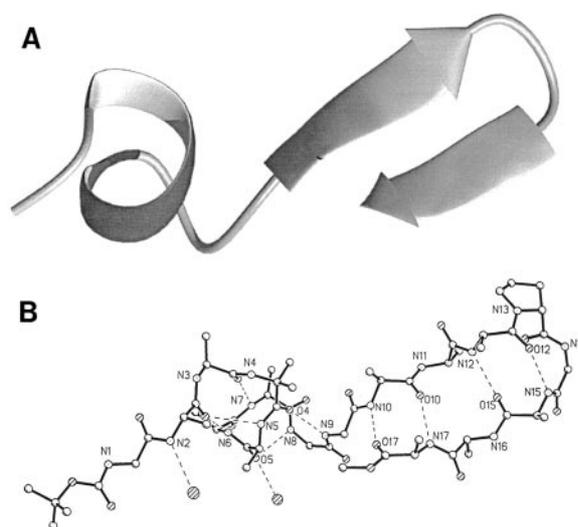
Isabella L. Karle<sup>\*†</sup>, Chittaranjan Das<sup>‡</sup>, and Padmanabhan Balaram<sup>‡</sup>

<sup>\*</sup>Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375-5341; and <sup>‡</sup>Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Contributed by Isabella L. Karle, January 31, 2000

The Meccano (or Lego) set approach to synthetic protein design envisages covalent assembly of prefabricated units of peptide secondary structure. Stereochemical control over peptide folding is achieved by incorporation of conformationally constrained residues like  $\alpha$ -aminoisobutyric acid (Aib) or  $\text{dPro}$  that nucleate helical and  $\beta$ -hairpin structures, respectively. The generation of a synthetic sequence containing both a helix and a hairpin is achieved in the peptide BH17, Boc-Val-Ala-Leu-Aib-Val-Ala-Leu-Gly-Gly-Leu-Phe-Val- $\text{dPro}$ -Gly-Leu-Phe-Val-OMe (where Boc is *t*-butoxycarbonyl), as demonstrated by a crystal structure determination. The achiral -Gly-Gly- linker permits helix termination as a Schellman motif and extension to the strand segment of the hairpin. Structure parameters for  $\text{C}_{89}\text{H}_{143}\text{N}_{17}\text{O}_{20}\cdot 2\text{H}_2\text{O}$  are space group  $\text{P}2_1$ ,  $a = 14.935(7)$  Å,  $b = 18.949(6)$  Å,  $c = 19.231(8)$  Å,  $\beta = 101.79(4)^\circ$ ,  $Z = 2$ , agreement factor  $R_1 = 8.50\%$  for 4,862 observed reflections  $> 4\sigma(F)$ , and resolution of  $\approx 0.98$  Å.

Several approaches to *de novo* protein design envisioning successful construction of protein-like structural motifs have been investigated over the past 2 decades (1). The Meccano (or Lego) set approach to synthetic protein design is based on the covalent assembly of prefabricated modules of peptide secondary structures (2, 3). Stereochemical control over peptide folding may be achieved by incorporation of conformationally constrained residues, like  $\alpha$ -aminoisobutyric acid (Aib) or  $\text{dPro}$ , that nucleate helical and  $\beta$ -hairpin structures, respectively (4). Nucleation of appropriate local conformations, together with the cooperative formation of several intramolecular hydrogen bonds in secondary structures, favors the population of stereochemically ordered segments in solvents that compete poorly for hydrogen bonding. The construction of helical segments of varying length has been achieved by using Aib residues for helix stabilization. Conformational characterization of peptide helices containing Aib has been documented extensively in a large number of single crystal x-ray diffraction studies (5–9). The use of  $\text{dPro}$ -Xxx segments to nucleate  $\beta$ -hairpin conformations is of more recent origin (10–13).  $\beta$ -Hairpins have been well characterized in solutions by using NMR spectroscopy, and the formation of three- and four-stranded  $\beta$ -sheets has been demonstrated in solution by incorporating multiple  $\text{dPro}$ -Xxx segments (14–18). Designed octapeptide  $\beta$ -hairpin modules Boc-L-V-V-DP-G-L-V-V-OMe (where Boc represents *t*-butoxycarbonyl and DP represents  $\text{dPro}$ ; ref. 19) and Boc-L-V-V-DP-A-L-V-V-OMe (I.L.K., P.B., and S. K. Awasthi, unpublished work) have been characterized in the crystalline state. The availability of peptide building blocks that adopt helical and  $\beta$ -hairpin conformations has enabled the construction of a synthetic 17-residue sequence, Boc-V-A-L-U-V-A-L-G-G-L-F-V-DP-G-L-F-V-OMe (BH17, where U represents Aib), that contains both elements of secondary structures linked together by a Gly-Gly segment (residues 8 and 9). The positioning of the achiral Gly residue at position



**Fig. 1.** (a) Ribbon drawing of BH17 showing the N-terminal extension, helix, linker, and  $\beta$ -hairpin (generated by the program MOLMOL; ref. 31). (b) Backbone in BH17 drawn with x-ray structure coordinates. Dashed lines indicate the intramolecular hydrogen bonds.

8 facilitates termination of the potential helical segment (residues 1–7) by formation of a Schellman motif (20–23). Gly-9 is anticipated to be the sole conformationally flexible linking residue. We describe the structure of BH17 in crystals, establishing an approximately orthogonal arrangement of the two elements of secondary structures, as shown in Fig. 1 *a* and *b*.

## Experimental Methods

The 17-residue peptide BH17 was synthesized by conventional solution phase procedures by using a fragment condensation strategy. Boc and methylester groups are used for N- and C-terminal protection. Peptide couplings were mediated by

Abbreviations: Boc, *t*-butoxycarbonyl; Aib,  $\alpha$ -aminoisobutyric acid or  $\alpha, \alpha'$ -dimethylglycine; V,  $\text{LVal}$ ; A,  $\text{LAla}$ ; G, Gly; L,  $\text{LLeu}$ ; F,  $\text{L Phe}$ ; U, Aib;  $\text{dP}$ ,  $\text{dPro}$ .

Data deposition: The atomic coordinates have been deposited in the Cambridge Crystallographic Data Centre, [www.ccdc.cam.ac.uk](http://www.ccdc.cam.ac.uk) (ID code CCDC 139326).

<sup>†</sup>To whom reprint requests should be addressed at: Laboratory for the Structure of Matter, Code 6030, Naval Research Laboratory, Washington, DC 20375-5341.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.070042697. Article and publication date are at [www.pnas.org/cgi/doi/10.1073/pnas.070042697](http://www.pnas.org/cgi/doi/10.1073/pnas.070042697)

**Table 1. Torsional angles in BH17**

Residue	Torsional angles, degrees							
	$\phi$	$\psi$	$\omega$	$\chi^1$	$\chi^2$	$\chi^3$	$\chi^4$	$\S$
Boc-0		178	-177					
Val-1	-129	+119	178	-58	-178			
Ala-2	-68	+148	-177					
Leu-3	-55	-52	-178	179	-178, +60			
Aib-4	-49	-44	-175					
Val-5	-71	-39	178	-63	175			
Ala-6	-60	-32	-176					
Leu-7	-97	+16	171	-65	177, -57			
Gly-8	+93	+11	-178					
Gly-9	+110	+175	-172					
Leu-10	-133	+121	167	-173	72, -166			
Phe-11	-72	+134	170	-69	-86	95		
Val-12	-119	+91	-166	177	-59			
DPro-13	+60	-131	178	+8	-28	+38	+16	-33
Gly-14	-82	-9	-179					
Leu-15	-92	171	-168	-62	172, -58			
Phe-16	-144	+126	-171	-172	-99, 83			
Val-17	-136	174	174	+73	-51			

Torsional angles  $\phi$ ,  $\psi$ , and  $\omega$  for the backbone and  $\chi^n$  for the side chains follow the convention presented in ref. 32. The estimated standard deviations are near 2.0°. The symbol  $\S$  represents the torsion angle  $C^{\alpha}NC^{\beta}$ .

*N,N'*-dicyclohexylcarbodiimide and 1-hydroxybenzotriazole. The N-terminal fragment (residues 1–7) and the C-terminal fragment (residues 8–17) were synthesized independently, followed by final extension to the target sequence. The peptide was purified by medium pressure liquid chromatography on a reverse phase  $C_{18}$  (10- to 60- $\mu$ m) column followed by HPLC on a  $C_{18}$  (5- to 10- $\mu$ m) column with methanol–water gradients.

Crystals of BH17 were obtained by slow evaporation from methanol containing a small amount of water. The crystals were very thin and fragile. Initially, x-ray diffraction data were collected at  $-50^{\circ}\text{C}$  from a crystal that was covered with microscope immersion oil immediately after removal from the mother liquor. The check reflections wandered in value during the data collection, and the number of data with observable intensity was somewhat limited. No structure solution was obtained. After 3 weeks of exposure to air, the same crystal yielded quite satisfactory data to a higher resolution. The x-ray data were measured at  $-60^{\circ}\text{C}$  on a four-circle diffractometer (Broker P4, Madison, WI) with  $\text{CuK}\alpha$  radiation ( $\lambda = 1.54178 \text{ \AA}$ ). The  $\theta/2\theta$  scan mode was used with a  $2.0^{\circ} + 2\theta(\alpha_1 - \alpha_2)$  scan width,  $15^{\circ}/\text{min}$  scan speed, and  $2\theta_{\text{max}} = 105^{\circ}$  ( $0.98\text{-\AA}$  resolution). The crystal size was  $0.07 \times 0.80 \times 0.90 \text{ mm}$ . The crystal data are  $\text{C}_{89}\text{H}_{143}\text{N}_{17}\text{O}_{20} \cdot 2\text{H}_2\text{O}$ , molecular weight =  $1,771.2 + 36$ , space group =  $\text{P}2_1$ ,  $a = 14.935(7) \text{ \AA}$ ,  $b = 18.949(6) \text{ \AA}$ ,  $c = 19.231(8) \text{ \AA}$ ,  $\beta = 101.79(4)^{\circ}$ ,  $V = 5,327.6 \text{ \AA}^3$ ,  $Z = 2$ , and calculated  $d = 1.124 \text{ g/cm}^3$ . The structure was solved with the aid of a vector search and translation procedure contained in the PATSEE program (24). The search model consisted of 23 backbone and  $\text{C}\beta$  atoms that are contained in the middle of the  $\alpha$ -helix in the acetyl carbonyl-V-A-L- $\alpha, \alpha'$ -dipropylglycine-V-A-L-OMe molecule (25). The best rotation and second-best translation position yielded 84 atoms when subjected to the tangent formula expansion procedure (26). The remainder of the atoms were found in difference maps interspersed with least-squares refinements. The final least-squares refinement—performed on all 7,148  $F^2$  values, with anisotropic thermal parameters, and 143 hydrogen atoms placed in idealized positions and riding with the C or N atoms to which they are bonded—resulted in  $R_1 = 8.50\%$  for 4,862 data with  $|F_o| > 4.0\sigma$  and 1,154 variables. (The least-squares program used was Siemens SHELXTL, version 5.03, Iselin, NJ.)

**Table 2. Hydrogen bonds**

Type	Donor	Acceptor**	Donor–acceptor, $\text{\AA}$	Hydrogen*–acceptor, $\text{\AA}$	CO...N angle, degrees
Intermol	N1	O11 <sup>†</sup>	3.035	2.17	143
Solvent	N2	W1	2.887	2.00	
Head-to-tail	N3	O9 <sup>†</sup>	2.870	2.01	147
Head-to-tail	N4	O7 <sup>†</sup>	2.979	2.14	169
Helix 4 $\rightarrow$ 1	N5	O2	2.975	2.41	123
Transition					
Helix 5 $\rightarrow$ 1	N6	O2	2.956	2.10	166
Helix 5 $\rightarrow$ 1	N7	O3	3.139	2.35	162
Helix 4 $\rightarrow$ 1	N8	O5	3.046	2.18	116
Helix reversal	N9	O4	2.818	2.02	150
$\beta$ -	N10	O17	3.001	2.16	158
Intermol	N11	O1 <sup>†</sup>	2.825	1.98	147
$\beta$ -	N12	O15	2.914	2.05	128
	N13 (DPro)				
Intermol	N14	O6 <sup>§</sup>	2.889	2.002	158
$\beta$ -	N15	O12	2.982	2.18	133
	N16	No hydrogen bond			
$\beta$ -	N17	O10	2.858	1.99	154
Solvent	W1	W2 <sup>¶</sup>	2.564		
Solvent	W2	Oo <sup>  </sup>	2.681		
Solvent	W2	O5	2.773		

\*Hydrogen atoms were placed in idealized positions with  $\text{N-H} = 0.90 \text{ \AA}$ .

<sup>†</sup>At symmetry equivalent  $-x, \frac{1}{2} + y, -z$ .

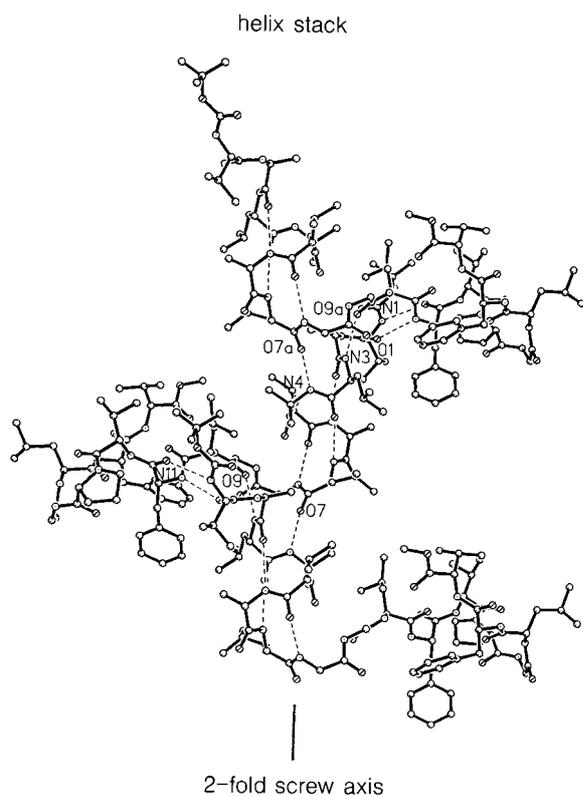
<sup>‡</sup>At symmetry equivalent  $-x, -\frac{1}{2} + y, -z$ .

<sup>§</sup>At symmetry equivalent  $-1 + x, y, -1 + z$ .

<sup>¶</sup>At symmetry equivalent  $1 - x, \frac{1}{2} + y, -z$ .

<sup>||</sup>At symmetry equivalent  $1 - x, -\frac{1}{2} + y, -z$ .

\*\*O8, O13, O14, and O16 do not participate in hydrogen bonding.



**Fig. 2.** Three molecules of BH17 are shown winding around the *b* axis (vertical). The helical portions are stacked vertically over each other and linked by N4...O7a-type hydrogen bonds.

## Results and Discussion

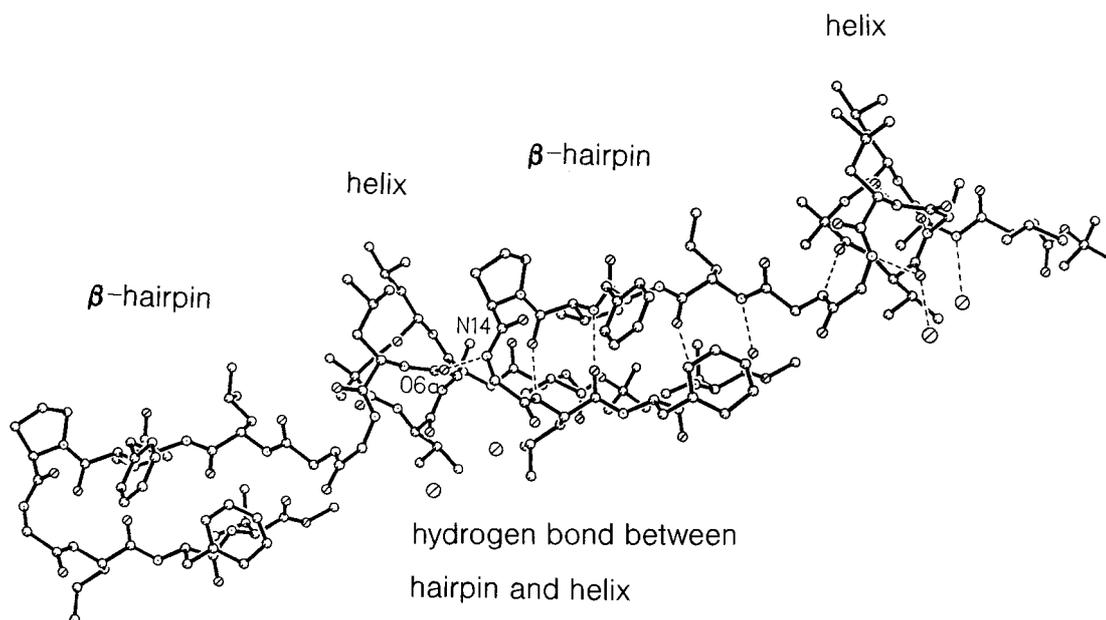
Inspection of the backbone torsion angles in Table 1 indicates that residues 3–7 form about one and a half turns of a helix. As

anticipated, Gly-8 adopts a left-handed helical ( $\alpha_L$ ) conformation, resulting in the formation of a Schellman motif, which is characterized by a 6  $\rightarrow$  1 hydrogen bond between Gly-9 NH and Aib-4 CO groups. Such a structural feature is a common mode of helix termination in proteins with Gly or Asn residues adopting the left-handed, helix-terminating conformation (20, 27, 28). Gly-9 is extended, allowing the polypeptide chain to progress comfortably into the N-terminal arm of the  $\beta$ -hairpin. Residues 10–17 form a classical  $\beta$ -hairpin nucleated by a type II'  $\beta$ -turn at the DPro-13–Gly-14 segment (19). The  $\beta$ -hairpin segment has nearly ideal intrastrand hydrogen bonds as shown in Table 2.

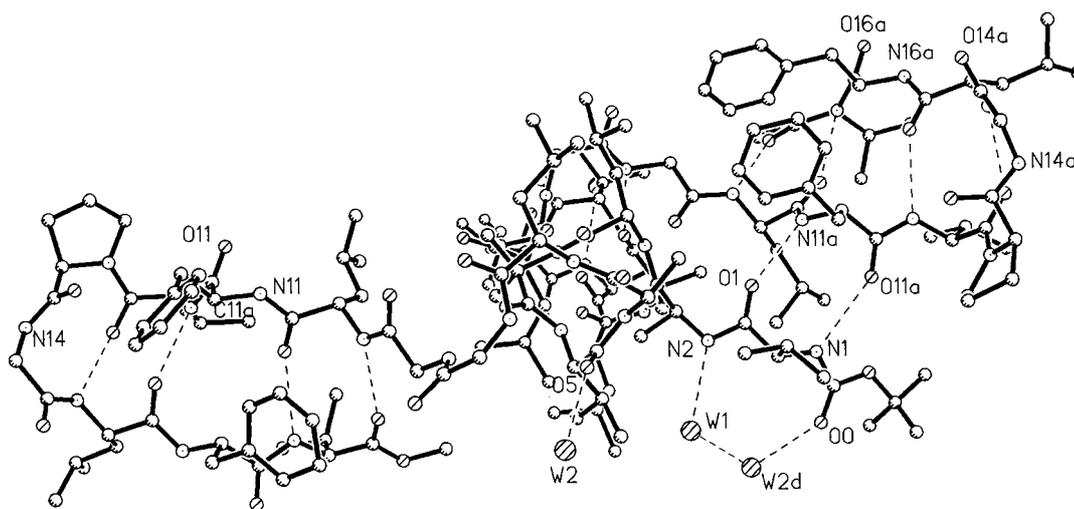
A striking feature of the peptide in crystals is the partially unfolded N-terminal helix, with Val-1 and Ala-2 adopting extended ( $\beta$ ) and semiextended conformations, respectively. Indeed, the heptapeptide module (residues 1–7) has been characterized as a helix in a large number of peptide crystal structures (29). However, in solution, this helix has been shown to be relatively fragile (30). Interestingly, in the present crystal structure, partial unfolding at the N terminus facilitates the formation of important intermolecular contacts.

There are several distinct features that characterize the assembly of these molecules in the crystal. One characteristic feature is the head-to-tail hydrogen bonding between the helical segments of adjacent molecules (N4...O7), even although the helical segments are not free-standing but are in the interior of the sequence (Fig. 2). The stacking of the helical segments simulates an infinite helix with a sinusoidal helix axis.

The two strands of the  $\beta$ -hairpin do not form an expected extended  $\beta$ -sheet by lateral intermolecular hydrogen bonding. N14 in the turn region of the  $\beta$ -hairpin forms a hydrogen bond with O6 in the helical portion of an adjacent molecule. The process is repeated by translation and forms an extended chain (Fig. 3). One arm of the  $\beta$ -hairpin is roughly antiparallel to the N-terminal extension of an adjacent molecule, as shown in Fig. 4, and a pair of hydrogen bonds are formed between the two structural elements (N1...O11a and N11a...O1). In this region, there is a limited extended  $\beta$ -sheet. Curiously, the other arm of the  $\beta$ -hairpin does not participate in any external hydrogen bonding. There is no acceptor for the N16H group. Atoms O14,



**Fig. 3.** Two molecules of BH17 related by translation. A continuous chain is formed by hydrogen bonds between N14 (at the turn of the  $\beta$ -hairpin) and O6 in the helix of an adjacent molecule.



**Fig. 4.** Two molecules of BH17 with their helical portions stacked in a direction perpendicular to the page and oriented such that the N-terminal extension and its accompanying hydrogen bonds are visible. On the right side is illustrated a limited, extended  $\beta$ -sheet formed by the two strands of the  $\beta$ -hairpin in one molecule and the N-terminal extension (Boc-Val-1-Ala-2) of an adjacent molecule.

O16, and N16H are directed into a hydrophobic pocket formed by methyl groups on the side chains of the helix and on the terminal methoxy group of several adjacent molecules. The closest approach is 3.35 Å between O16 and C3D2 of Leu-3.

Two water molecules, W1 and W2, occupy a polar pocket along the extended N terminus on the other side from the  $\beta$ -hairpin (Fig. 4). W1 forms hydrogen bonds with N2 and W2, whereas W2 forms additional hydrogen bonds to O0 and O5 (in a symmetry-related molecule).

## Conclusions

The crystal structure of the peptide BH17 validates the assumption that the structural domains within the molecule are indeed independent, suggesting that a modular assembly strategy may be viable for the construction of larger synthetic structures.

This research was partially supported in Bangalore by the program on Drug and Molecular Design of the Department of Biotechnology, Government of India, as well as by National Institutes of Health Grant GM30902 and by the Office of Naval Research.

- DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. (1999) *Annu. Rev. Biochem.* **68**, 779–819.
- Balaram, P. (1992) *Pure Appl. Chem.* **64**, 1061–1066.
- Balaram, P. (1999) *J. Peptide Res.* **54**, 195–199.
- Kaul, R. K. & Balaram, P. (1999) *Bioorg. Med. Chem.* **7**, 105–117.
- Balaram, P. (1992) *Curr. Opin. Struct. Biol.* **2**, 845–851.
- Karle, I. L. (1999) *Acc. Chem. Res.* **32**, 693–701.
- Karle, I. L. & Balaram, P. (1990) *Biochemistry* **29**, 6747–6756.
- Toniolo, C. & Benedetti, E. (1988) *ISI Atlas Sci. Biochem.* **1**, 225–230.
- Toniolo, C. & Benedetti, E. (1991) *Trends Biochem. Sci.* **16**, 350–353.
- Awasthi, S. K., Raghobama, S. & Balaram, P. (1995) *Biochem. Biophys. Res. Commun.* **216**, 375–381.
- Haque, T. S., Little, J. C. & Gellman, S. H. (1996) *J. Am. Chem. Soc.* **118**, 6975–6985.
- Haque, T. S. & Gellman, S. H. (1997) *J. Am. Chem. Soc.* **119**, 2303–2304.
- Struthers, M. D., Cheng, R. P. & Imperiali, B. (1996) *Science* **271**, 342–345.
- Lacroix, E., Kortemme, T., de la Paz, M. L. & Serrano, L. (1999) *Curr. Opin. Struct. Biol.* **9**, 487–493.
- Raghobama, S. R., Awasthi, S. K. & Balaram, P. (1998) *J. Chem. Soc. Perkin Trans. 2*, 137–143.
- Das, C., Raghobama, S. & Balaram, P. (1998) *J. Am. Chem. Soc.* **120**, 5812–5813.
- Schenck, H. L. & Gellman, S. H. (1998) *J. Am. Chem. Soc.* **120**, 4869–4870.
- Das, C., Raghobama, S. R. & Balaram, P. (1999) *J. Chem. Soc. Chem. Commun.* 967–968.
- Karle, I. L., Awasthi, S. K. & Balaram, P. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8189–8193.
- Gunasekaran, K., Nagarajaram, H. A., Ramakrishnan, C. & Balaram, P. (1998) *J. Mol. Biol.* **215**, 917–932.
- Karle, I. L., Flippen-Anderson, J. L., Uma, K. & Balaram, P. (1993) *Int. J. Peptide Protein Res.* **42**, 401–410.
- Datta, S., Shamala, N., Banerjee, A., Pramanik, A., Bhattacharjya, S. & Balaram, P. (1997) *J. Am. Chem. Soc.* **119**, 9246–9251.
- Banerjee, A., Datta, S., Pramanik, A., Shamala, N. & Balaram, P. (1996) *J. Am. Chem. Soc.* **118**, 9477–9483.
- Egert, E. & Sheldrick, G. M. (1985) *Acta Crystallogr. A* **41**, 262–268.
- Vijayalakshmi, S., Rao, R. B., Karle, I. L. & Balaram, P. (2000) *Biopolymers* **53**, 84–98.
- Karle, J. (1968) *Acta Crystallogr. B* **24**, 182–186.
- Schellman, C. (1980) in *Protein Folding*, ed. Jaenicke, R. (Elsevier, North-Holland, Amsterdam), pp. 53–61.
- Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 164–339.
- Karle, I. L., Gurunath, R., Prasad, S., Kaul, R., Rao, R. B. & Balaram, P. (1995) *J. Am. Chem. Soc.* **117**, 9632–9637.
- Karle, I. L., Flippen-Anderson, J. L., Uma, K. & Balaram, P. (1993) *Biopolymers* **33**, 827–837.
- Koradi, R., Billeter, M. & Wuthrich, K. (1996) *J. Mol. Graphics* **14**, 51–55.
- IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9**, 3471–3479.