Pure & Appl. Chem., Vol. 64, No. 8, pp. 1061–1066, 1992.
Printed In Great Britain.

The design and construction of synthetic protein mimics

P. Balaram

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

Abstract - A strategy for the modular construction of synthetic protein mimics based on the ability of non-protein amino acids to act as stereochemical directors of polypeptide chain folding, is described. The use of α-aminoisobutyric acid (Aib) to construct stereochemically rigid helices is exemplified by crystallographic and spectroscopic studies of several apolar peptides, ranging in length from seven to sixteen residues. The problem of linker design in elaborating α,α motifs is considered. Analysis of protein crystal structure data provides a guide to choosing linking sequences. Attempts at constructing linked helical motifs using linking Gly-Pro segments are described. Ths use of flexible linkers, like ε-aminocaproic acid is examined and the crystallographic and solution state analysis of a linked helix motif is presented. The use of bulky sidechain modifications on a helical scaffold, as a means of generating putative binding sites is exemplified by a crystal structure of a peptide packed in a parallel zipper arrangement.

INTRODUCTION

The synthetic construction of peptides that mimic the supersecondary structural motifs in proteins requires that suitably designed sequences fold in predictable fashion (ref. 1). Polypeptide sequences employing the twenty genetically coded amino acids possess a degree of structural flexibility that makes definitive predictions difficult, especially for sequences that have been designed de novo (ref. 2,3). An approach that is being developed in this laboratory employs non-protein amino acids as stereochemical directors of polypeptide chain folding (ref. 4,5). In this strategy conformationally rigid modules of secondary structure, like helices, are constructed from sequences containing amino acids with overwhelmingly strong stereochemical preferences. For instance, α-aminoisobutyric acid (Aib), a common constituent of many fungal peptide antibiotics (ref. 6), facilitates helix formation in oligopeptides and stabilises 3_{10}/α-helical structures in relatively short peptides (ref. 6,7). The availability of helical modules then permits further elaboration of sequences containing linked, secondary structure elements. This strategy for construction of synthetic protein mimics is schematically illustrated in Fig. 1. The key steps in this approach are:

(i) Choice of appropriate structural motif.
(ii) Design, synthesis and characterisation of structured modules.
(iii) Design of linking segment sequences.
(iv) Synthesis and structural characterisation of protein mimics,

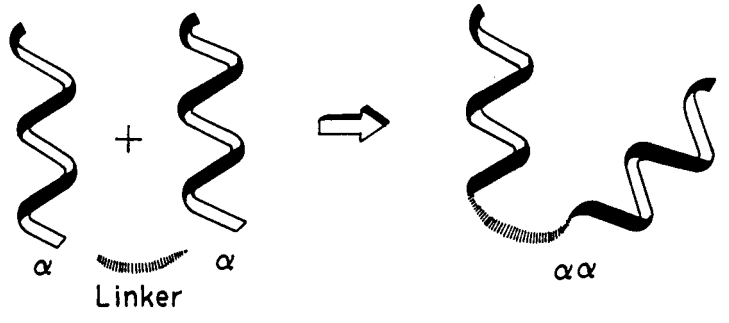

Fig. 1.
A modular `Meccano set' approach to synthetic protein design. Schematic illustration of α,α -motif construction.

NON-PROTEIN AMINO ACIDS IN PEPTIDE DESIGN

Figure 2 lists the structures of some non-protein amino acid residues that are being investigated. Aib (ref. 6,7) and the 1-aminocycloalkane-1-carboxylic acid (Acnc) (ref. 8,9) residues strongly favour heical conformations, with backbone dihedral angles of $\phi \sim \pm 50^{0} \pm 30^{0}$, $\psi \sim \pm 50^{0} \pm 30^{0}$. Indeed in the case of Aib, even incorporation of a single residue into a heptapeptide (ref. 10) or hexapeptide (ref. 11) results in the stabilisation of helical conformations in crystals. A particularly dramatic example is the peptide Boc-Val-Val-Aib-Pro-Val-Val-Val-OMe, where a 3_{10}-helix is observed in the crystalline state, even though the central Pro residue interrupts a regular chain of intramolecular hydrogen bonds (ref. 11). Similarly, an overall helical fold

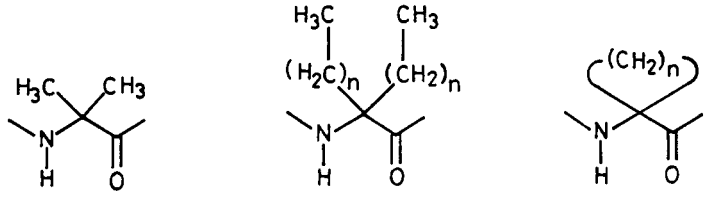

Fig. 2. Non-protein amino acid residues. Aib, α-aminoisobutyric acid; Deg, diethylglycine; Dpg, dipropylglycine; Ac_5c, 1-aminocyclopentane carboxylic acid; Ac_6c, 1-aminocyclohexane carboxylic acid.

is maintained in the C-terminal segment of the fungal peptide zervamicin (ref. 12) and a synthetic analog (ref. 13), despite the presence of as many as three Pro/Hyp residues. Aib residues can therefore be used to promote helical structures, a property that has been used to advantage in extensive studies of peptide helices in crystals and solution. The structural characterisation, at high resolution < 1.0 Å, of helical peptides in single crystals has afforded many new insights into helix packing, hydration and conformational heterogeneity in crystals (ref. 7). Helices ranging in length from 7 to 16 residues have been characterised in apolar sequences containing Aib residues. Structures which incorporate 3 to 4 helical turns in their crystal state conformations include the peptides Boc-(V-A-L-U)$_4$-OMe (ref. 14), Boc-V-A-L-U-V-A-L-V-A-L-U-V-A-L-U-OMe (ref. 14) and Boc-U-(V-A-L-U)$_3$-OMe (ref. 15). (Single letter amino acid code: V= Val, A= Ala, L= Leu, U= Aib). CD spectra characteristic of highly helical sequences are obtained in methanol solutions, suggesting that solid state and solution conformations are similar (ref. 16).

Far fewer studies have been carried out on the $C^{\alpha\alpha}$-dialkylated amino acids containing longer alkyl sidechains like diethylglycine (Deg), dipropylglycine (Dpg) and dibutylglycine (Dbg). Studies of homooligomers in crystals and theoretical calculations have suggested that Deg, Dpg and Dbg favour fully extended (C_5) conformations. Preliminary investigations of tripeptides of the type Boc-Ala-X-Ala-OMe have shown that for X = Dpg, Dbg extended forms are favoured, whereas for X = Acnc (n = 6,7), β-turn or folded structures are preferred (ref. 17-20). While characterisation of longer sequences containing Dpg and Dbg are awaited, it is clear that these residues may be useful in forcing polypeptide chains to extended forms.

LINKING HELICES

While stereochemically rigid modules like helices and disulfide bridged antiparallel hairpins (ref. 21) may be constructed, further assembly requires the design of linking peptide sequences that are necessary to connect individual structural elements. Analysis of the amino acid compositions and backbone conformations of short linking loops in crystalline proteins, provides a means of designing linkers.

Using a 65 protein data set, all $\alpha\alpha$, $\beta\beta$, $\alpha\beta$ and $\beta\alpha$ motifs containing 1 to 5 residue linking segments have been examined. Fig. 3 illustrates results of such an analysis for $\alpha\alpha$ motifs (ref. 22). The identification of linkers that fall into well defined conformational families

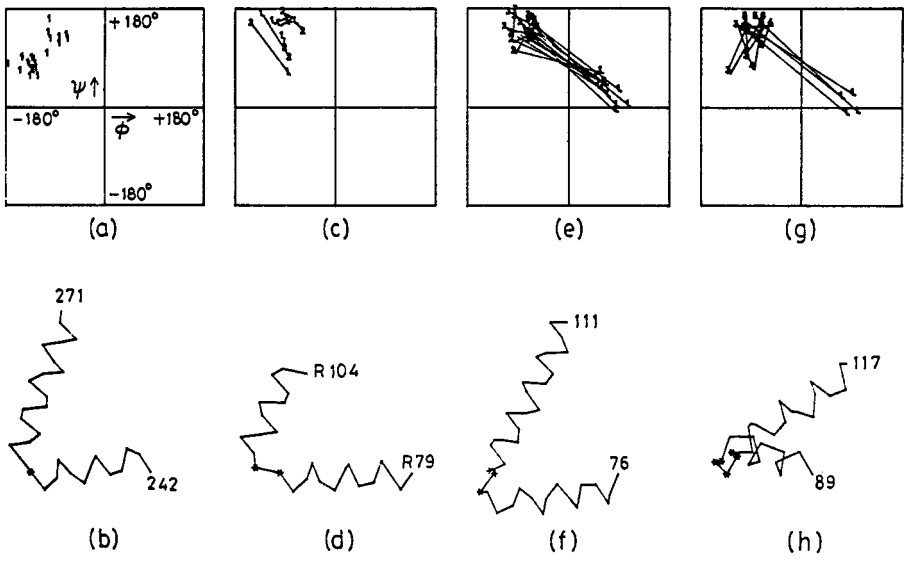

Fig.3. Conformational families observed for linking loops in α,α motifs, represented on a (ϕ,ψ) plot. C^{α} tracings of representative motifs with residue numbers indicated are shown below each plot. The numbers 1-5 in the (ϕ,ψ) diagrams mark the conformations at residues 1-5 of the linking segment. Residues in the linking loops are marked with an asterix in the C^{α} tracings. (a) single residue linkers, (b) 242-271 cytochrome c peroxidase (2CYP), (c) two residue linkers, (d) R79-R104 trp-repressor (2WRP), (e) three residue linkers, (f) 76-111 erythrocuprorin (1ECA), (g) five residue linkers, (h) 89-117 citrate synthase (2CTS) (from ref. 22).

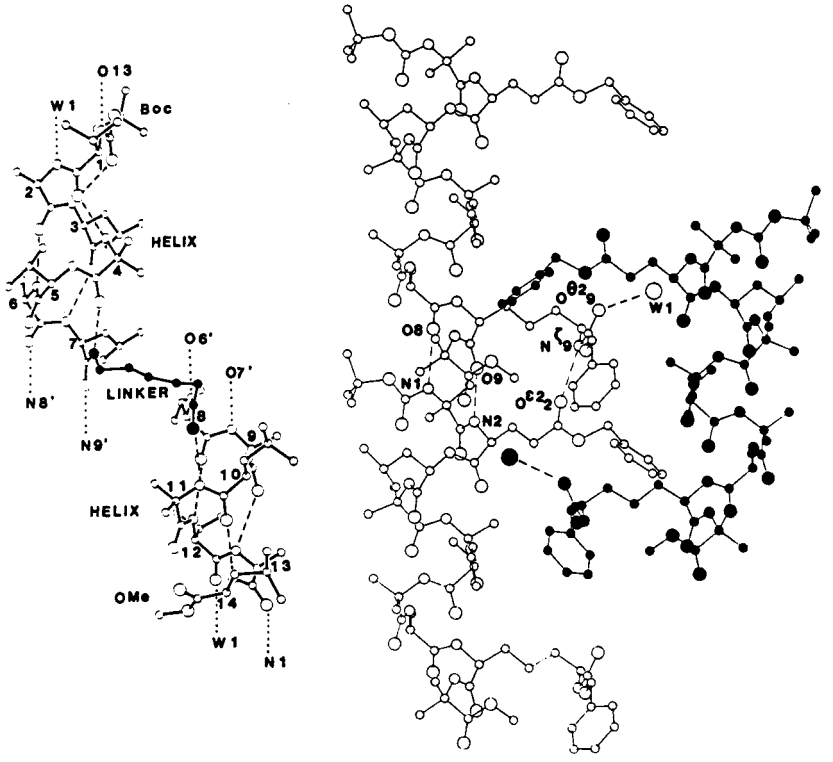

Fig. 4. (left) Crystal state conformation of Boc-Val-Ala-Leu-Aib-Val-Ala-Leu-Acp-Val-Ala-Leu-Aib-Val-Ala-Leu-OMe, illustrating displaced helical modules (from ref. 25), (right) Crystal structure of Boc-Aib-Glu(OBzl)-Leu-Aib-Ala-Leu-Aib-Ala-Lys(Z)-Aib-OMe illustrating insertion of sidechains from adjacent columns into a `binding site' cavity on a neighbour (marked in black) (from ref. 27).

suggests that one approach to the problem of interhelix orientation would be to use stereochemically constrained amino acids with a strong preference for the appropriate region of ϕ, ψ space (ref. 5). This would be particularly important in design of motifs, where interhelix contacts are minimal as in the case of nearly orthogonal, $\alpha\alpha$ motifs (ref. 22, 23). The results also show a clear preference of some amino acids like Gly for the linking segments. Initial attempts in this laboratory to use a central Gly-Pro segment flanked by two well characterised helices resulted in a 18-residue peptide Boc-Aib-Val-Ala-Leu-Aib-Val-Ala-Leu-Gly-Pro-Val-Ala-Leu-Aib-Val-Ala-Leu-Aib-OMe. This peptide failed to show an appreciable break in the helical folding pattern as judged by 2D-NMR, CD and molecular dimensions obtained from a determination of cell parameters of single crystals in the space group $P4_1$ (ref. 16, 24). Indeed all data were consistent with a largely continuous helix, despite the presence of 8 contiguous non-Aib residues, including the potentially helix breaking Gly-Pro unit. Ironically, this result suggests that breaking a long helical segment nucleated by even a couple of Aib residues may not be an easy proposition. A straight forward approach to this problem would be to enhance the flexibility of the linker. This was achieved by using ε-aminocaproic acid (Acp) as a linker in the peptide Boc-Val-Ala-Leu-Aib-Val-Ala-Leu-Acp-Val-Ala-Leu-Aib-Val-Ala-Leu-OMe (1). The conformation in crystals is shown in Fig. 4 (ref. 25). It is clearly seen that both heptapeptide segments retain their helical conformations. However, the axes of the two cylindrical helical modules are displaced by the Acp residue although the two individual helices remain approximately parallel, but extended. The central CO (Ala(6), Leu(7)) and NH (Val(9), Ala(10)) groups which do not form intramolecular hydrogen bonds interact intermolecularly with corresponding donor and acceptor groups on adjacent molecules, stabilising the conformation observed in crystals. Interestingly, the HPLC retention time of peptide 1 on a reverse phase C_{18} column is appreciably lower than that of single continuous helices of comparable size. Fig. 5

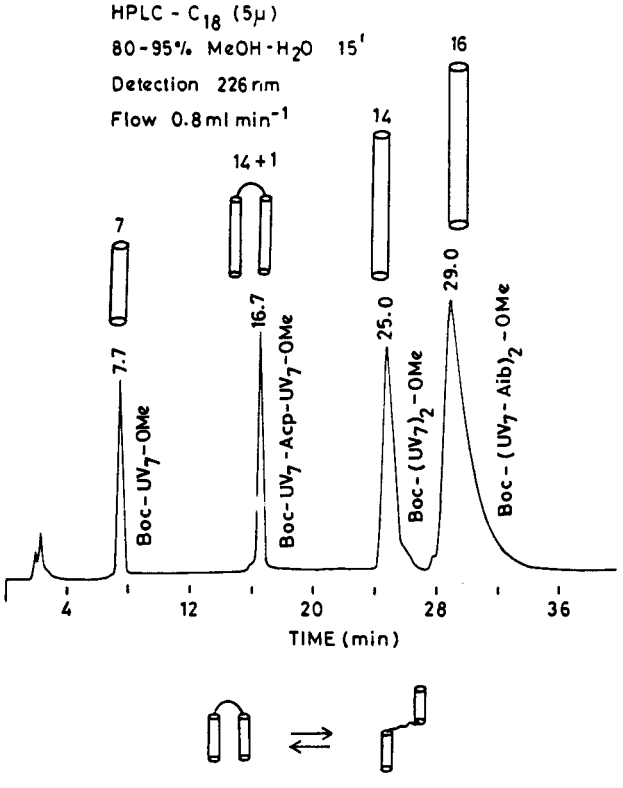

Fig. 5. HPLC trace of a peptide mixture of varying length, indicated above the schematic cylindrical structures. Retention times marked above the peaks. Peptide abbreviation used is UV_7 = -Val-Ala-Leu-Aib-Val-Ala-Leu-. Interconversions between antiparallel, close packed cylinders and open, extended forms is illustrated at the bottom.

shows a chromatogram of a peptide mixture of helical peptides, of varying length and similar sequence, whose rodlike helical conformations have been established in both the solid state and solution. Peptide 1 has a dramatically lower retention time as compared to cylindrical molecules with nearly the same number of residues. This suggests that the exposed surface area, accessible for interactions with the C_{18} chains on the column, is appreciably lower. This observation favours a compact conformation in which the two helical modules close pack in antiparallel fashion as shown schematically in Fig. 5. The packing of large complementary surfaces in organic solvents can indeed be driven by solvophobic interactions (ref. 26). Modelling studies indicate that rotation about the single bonds of the flexible pentamethylene chain of the Acp linker suffices to reorient the helices comfortably. Indeed, the present results suggest a degree of structural flexibility in 1 with the orientation of the helical modules being determined by environmental factors. A 31-residue peptide containing four helical modules and three linking Acp residues Boc-(Val-Ala-Leu-Aib-Val-Ala-Leu-Acp)$_3$-Val-Ala-Leu-Aib-Val-Ala-Leu-OMe has been synthesised and conformational characterisation is in progress.

The control of helix orientation may be achieved by building in electrostatic intereactions between the helical modules. As a first attempt in this direction, the 18-residue peptide Boc-Aib-Ala-Aib-Leu-Aib-Lys-Val-Leu-Gly-Pro-Asp-Ala-Leu-Aib-Ala-Aib-Leu-Aib-OMe was designed with a central Gly-Pro linking unit. A comparison of CD spectra of both the charged peptide and the protected analog (Lys (Z) and Glu (OBzl)) revealed very similar helical conformations. Indeed, the CD data were consistent with a continuous helical conformation of both molecules. During the course of studies aimed at building up helical modules with a potential for charge introduction, the sidechain protected peptide Boc-Aib-Glu(OBzl)-Leu-Aib-Ala-Leu-Aib-Ala-Lys(Z)-Aib-OMe was synthesised and crystallised. The crystal structure (Fig. 4) revealed a perfectly helical conformation for the molecule with a novel packing arrangement (ref. 27). The bulky protected sidechains of Lys and Glu project perpendicular to the helix axis resulting in the formation of a large cup-like, acyclic cavity. Portions of adjacent helical columns intrude into this `putative binding site'. The arrangement of helices, head-to-tail within columns which are parallely packed, gives rise to a novel zipper like association of adjacent columns. This structure suggests that the residue separation of the bulky sidechains is optimal for creating a potential `host' site. Indeed, a heptad repeat of the smaller isobutyl sidechains of Leu residues on a helical backbone is a feature of the `knobs in holes' packing (ref. 28) envisaged for leucine zippers (ref. 29) and coiled coil proteins (ref. 30). The structure described above suggests that construction of orthogonal peptides by means of functionalised sidechains could lead to new molecules with interesting binding properties. Ongoing studies in this laboratory are aimed at exploring such structures. The present approach aims at developing a well organised molecular scaffold, using defined polypeptide chain folding patterns, as a means of controlling spatial orientation of reactive sidechains. The use of hydrophobic helical modules results in relatively high solubility in organic solvents, of peptides of considerable length. This facilitates studies of folding under conditions where the dominant interactions can be modelled and hydrophobic effects are absent. The construction of amphipathic sequences should permit extension to aqueous solutions with inter-module association being driven by hydrophobic interactions.

Acknowledgements I am deeply indebted to Dr. Isabella Karle (Naval Research Laboratory, Washington D.C. U.S.A.) for her continued collaboration in this project, and Dr. K. Uma and Dr. M. Sukumar for their extensive studies over several years. The protein data analysis has been carried out in collaboration with R. Sowdhamini, Dr. N. Srinivasan and Dr. C. Ramakrishnan. Financial support from the Department of Science and Technology, Government of India, is gratefully acknowledged.

REFERENCES

1. W.F. Degrado, Adv. Protein Chem. 39, 51-124 (1988).
2. M.H. Hecht, J.S. Richardson, D.C. Richardson and R.C. Ogden, Science, 249, 884-891 (1990).
3. W.F. DeGrado, Z.R. Wasserman and J.D. Lear, Science 243, 622-628 (1989).
4. P. Balaram, Proc. Ind. Acad. Sci. Chem. Sci. 93, 703-717 (1984).

5. I.L. Karle, J.L. Flippen-Anderson, K. Uma and P. Balaram, Curr. Sci. 59, 975-985 (1990).

6. B.V.V. Prasad and P. Balaram, CRC Crit. Rev. Biochem. 16, 307-347 (1984).

7. I.L. Karle and P. Balaram, Biochemistry, 29, 6747-6756 (1990).

8. P.K.C. Paul, M. Sukumar, R. Bardi, A.M. Piazzesi, G. Valle, C. Toniolo and P. Balaram, J. Am. Chem. Soc. 108, 6363-6370 (1986).

9. G. Valle, M. Crisma, C. Toniolo, Sudhanand, R. Balaji Rao, M. Sukumar and P. Balaram, Int. J. Peptide Protein Res. in press (1991)

10. I.L. Karle, J.L. Flippen-Anderson, K. Uma and P. Balaram, Proteins: Struct. Funct. Genetics 7, 62-73 (1990).

11. I.L. Karle, J.L. Flippen-Anderson, K. Uma, H. Balaram and P. Balaram, Biopolymers, 29, 1433-1442 (1990).

12. I.L. Karle, J.L. Flippen-Anderson, S. Agarwalla and P. Balaram, Proc. Natl. Acad. Sci. U.S.A. 88, 5307-5311 (1991).

13. I.L. Karle, J.L. Flippen-Anderson, M. Sukumar and P. Balaram, Proc. Natl. Acad. Sci. U.S.A. 84, 5087-5091 (1987).

14. I.L. Karle, J.L. Flippen-Anderson, K. Uma, M. Sukumar and P. Balaram, J. Am. Chem. Soc. 112, 9350-9356 (1990).

15. I.L. Karle, J.L. Flippen-Anderson, K. Uma and P. Balaram, Biochemistry 28, 6696-6701 (1989).

16. K. Uma, Modular Design of Synthetic Protein Mimics. Construction of Helices. Ph.D. Thesis Indian Institute of Science, 1990. Thesis Abstract. J. Ind. Inst. Sci. 71 395-398 (1991).

17. C. Toniolo and E. Benedetti, ISI Atlas of Science Biochem. 1, 225-230 (1988).

18. C. Toniolo, British Polymer J. 18 221-225 (1986).

19. C. Toniolo, G.M. Bonora, A. Bavoso, E. Benedetti, B. DiBlasio, V. Pavone, C. Pedone, V. Barone, F. Lelj, M.T. Leplawy, K. Kaczmarek and A. Redlinski, Biopolymers, 27 373-379 (1988).

20. Sudhanand, R. Balaji Rao and P. Balaram, manuscript in preparation

21. I.L. Karle, J.L. Flippen-Anderson, R. Kishore, S. Raghothama and P. Balaram, J. Am. Chem. Soc. 110, 1958-1963 (1988).

22. N. Srinivasan, R. Sowdhamini, C. Ramakrishnan and P. Balaram, Molecular Conformation and Biological Interactions (Eds. P. Balaram and S. Ramaseshan), Indian Academy of Sciences, in press (1991).

23. A.V. Efimov, FEBS Letters, 166, 33-38 (1984).

24. K. Uma, I.L. Karle and P. Balaram, in 'Proteins: Structure, Dynamics and Design', (Eds, V. Renugopalakrishnan, P. R. Carey, I.C.P. Smith, S.G. Huang and A. Storer), Escom Science Publishers B.V. Leiden, pp 295-301 (1991).

25. I.L. Karle, J.L. Flippen-Anderson, M. Sukumar, K. Uma and P. Balaram, J. Am. Chem. Soc. 113, 3952-3956 (1991).

26. J.A. Bryant, C.B. Knobler and D.J. Cram, J. Am. Chem. Soc. 112, 1254-1255 (1990).

27. I.L. Karle, J.L. Flippen-Anderson, K. Uma and P. Balaram, Proc. Natl. Acad. Sci. U.S.A. 87, 7921-7925 (1990).

28. F.H.C. Crick, Acta Crystallogr. 6, 689-697 (1953).

29. C.R. Vinson, P.B. Sigler and S.L. McKnight, Science 246, 911-916.

30. C. Cohen and D.A.D. Parry, Proteins: Struct. Funct. Genetics, 7, 1-15 (1990).