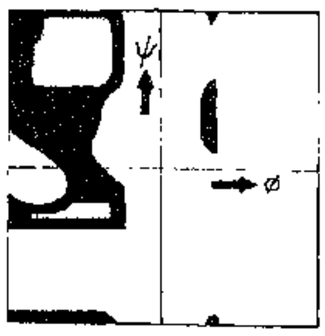

Special issue:
The Ramachandran map and
protein and peptide structures

Helix construction using α-aminoisobutyryl residues in a modular approach to synthetic protein design. Conformational properties of an apolar decapeptide in two different crystal forms and in solution

Isabella L. Karle[a]*, J. L. Flippen-Anderson[a], K. Uma[b] and P. Balaram[b]*

[a] Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375-5000, USA
[b] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

A modular strategy for the assembly of synthetic protein mimics is outlined. This approach is based on the ability of α-aminoisobutyric acid (Aib) to promote helical folding in peptides. The design of apolar, peptide helices with a high solubility in organic solvents facilitates structural design based on stereochemical constraints imposed by chosen non-protein residues. The design of individual helices is illustrated by the analysis of the decapeptide Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe. Crystal structures of two polymorphic forms reveal an almost completely α-helical backbone. These high-resolution structures permit detailed characterization of the helix. ^{1}H NMR and CD studies demonstrate maintenance of the helical conformation in solution, a feature that is a consequence of the limited flexibility of Aib residues in ϕ, ψ space.

The three-dimensional folding patterns of the polypeptide chain in proteins can be conveniently analysed in two-dimensional torsion angle (ϕ, ψ) space by means of Ramachandran maps[1,2]. The backbone stereochemistry of globular proteins is determined by the values of the ϕ, ψ torsion angles at each residue, with the overwhelming majority of peptide units adopting the rigid *trans* ($\omega = 180°$) geometry[3]. One of the aims of contemporary peptide and protein research is to develop strategies for the construction of synthetic proteins with novel properties[4–11]. The difficulties of predicting backbone folding patterns from primary sequence information[12,13] are major obstacles in the path of *de novo* protein design. An approach which is being developed in these laboratories attempts to exploit the novel structural properties of unusual, non-protein amino acids in designing oligopeptides of defined stereochemistry[14,15]. The use of conformationally rigid residues should permit the construction of structurally well-defined modules, which can then be assembled by flexible 'linker peptides' into supersecondary structural motifs as illustrated in Figure 1.

The success of this modular approach to construction of synthetic protein mimics, which has been termed as a 'Meccano set' strategy[15] depends on several factors:

(i) The availability of amino acid residues which strongly favour specific regions of the Ramachandran map and act as structure nucleating/stabilizing elements. Expanding the repertoire of conformationally constrained amino acids is central to this approach.

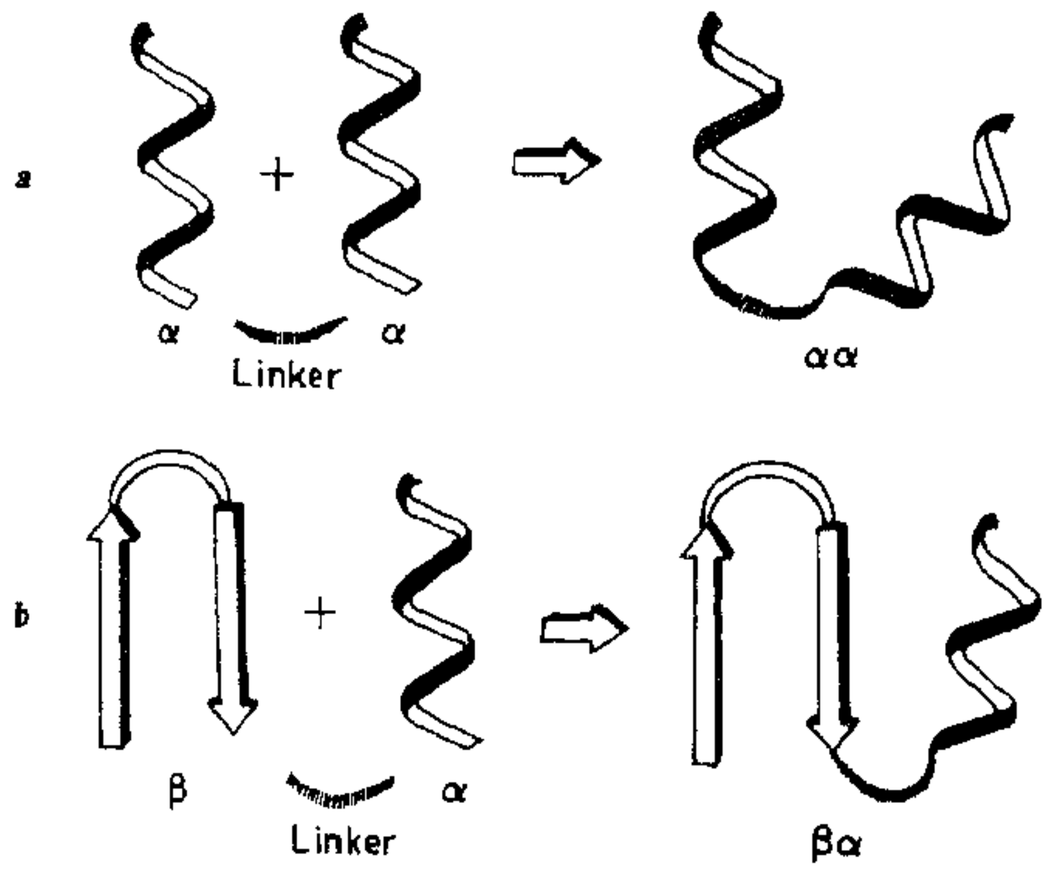

Figure 1. Schematic illustration of the modular or 'Meccano set' approach to the construction of (*a*) α, α and (*b*) β, α motifs.

*For correspondence.

(ii) The ability to use such conformationally rigid residues to construct small units of secondary structure, like α-helices, extended strands and β-turns, which will then serve as modules for further synthesis.

(iii) The proper choice of linking sequences, which will permit appropriate orientation of the designed structured modules. This may be accomplished by analysis of the compositional and conformational preferences of linking loop segments in protein crystal structures[16].

(iv) The design of sequences which retain a high solubility in organic solvents, so that folding can be driven in a relatively easily predictable way, by van der Waals interactions, hydrogen bonding and electrostatic effects, eliminating hydrophobic interactions as a determinant of the final structure.

The most extensively investigated group of conformationally constrained residues are the α,α-dialkylated amino acids (Figure 2)[17]. Conformational energy calculations[18] first suggested that the α-aminoisobutyric acid (Aib) residue was likely to favour helical conformations due to the occurrence of deep minima in this region of ϕ,ψ space. The ability of Aib residues to stabilize folded structures in the solid state and solution was demonstrated by the observation of incipient 3_{10}-helical structures in a synthetic tetrapeptide[19,20] and a pentapeptide[21]. Following these early studies, a large body of experimental evidence has been accumulated for the ability of Aib residues to stabilize helical structures[17,22–26]. The conformational analysis of Aib-containing peptides has largely been stimulated by the widespread occurrence of this residue in fungal peptides, which form voltage-sensitive channels across lipid bilayer membranes[27–32].

Figure 3 demonstrates that the overwhelming majority of crystallographically determined ϕ, ψ values for the Aib residue in crystalline peptides lie in the left and right handed α-helical regions of conformational space. These results provide a strong basis for investigating the use of Aib residues in the design of synthetic helix modules for the assembly of protein mimics. We have been systematically investigating the

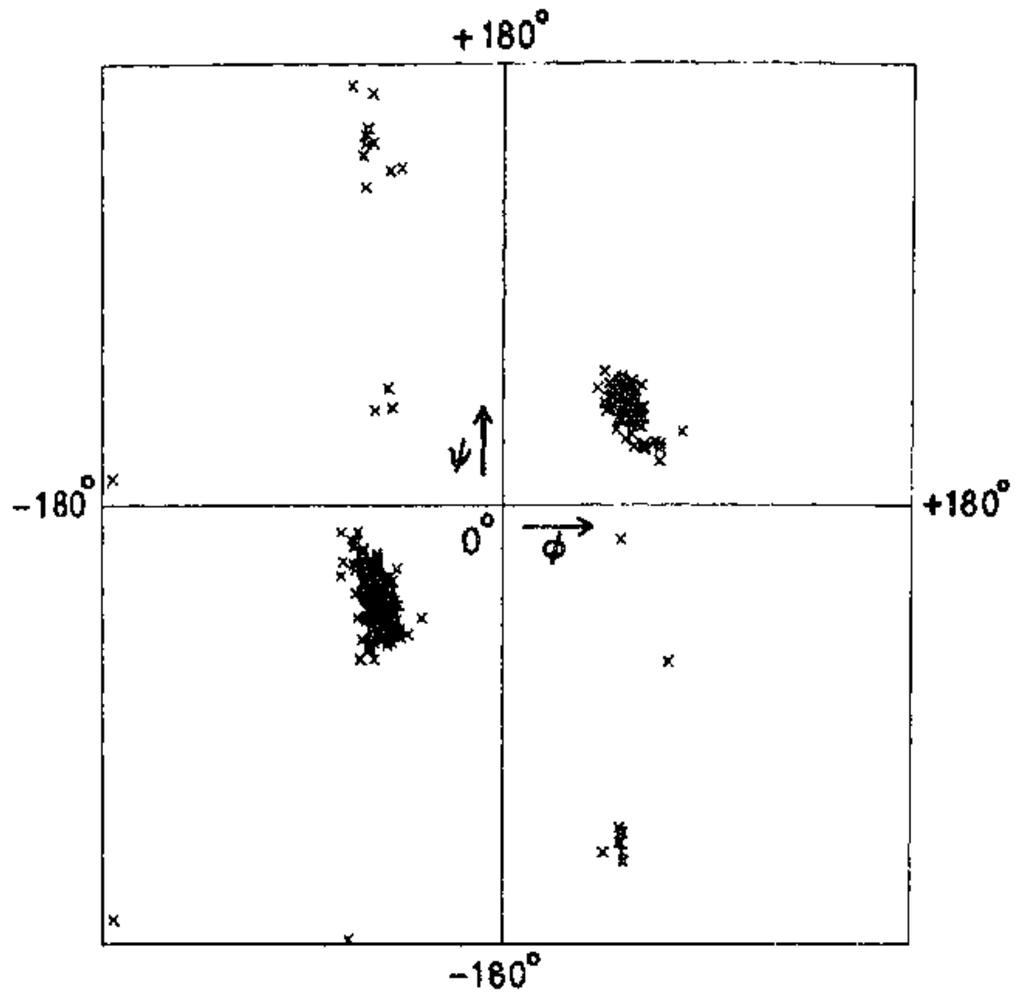

Figure 3. Crystallographically observed ϕ, ψ values (×) of 301 Aib residues from 105 independent crystal structures. In the case of achiral peptides crystallizing in centrosymmetric space groups, the sign of the dihedral angles has been chosen arbitrarily.

role of Aib content and position on the structural integrity of helices varying in length from 9 to 16 residues. The crystallizability of Aib-containing peptides has permitted the unambiguous characterization of several helical structures by X-ray diffraction[15,33–46]. The role of packing effects in determining the detailed structure of these peptides and the stability of the helices in different environmental conditions are being investigated by studying several crystalline polymorphs[39,40]. The different modes of helix aggregation in separate crystal systems[33,39,40], the incorporation of solvent molecules between helices[15,35,37,40,44], backbone solvation[34,37], conformational heterogeneity[36,37], and the occurrence of voids[33,44] are of special interest. The structure of the decapeptide Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe (I) crystallized from isopropanol is examined in this paper and compared with a polymorph obtained from methanol[45]. The structural characteristics of related peptide helices are also briefly compared.

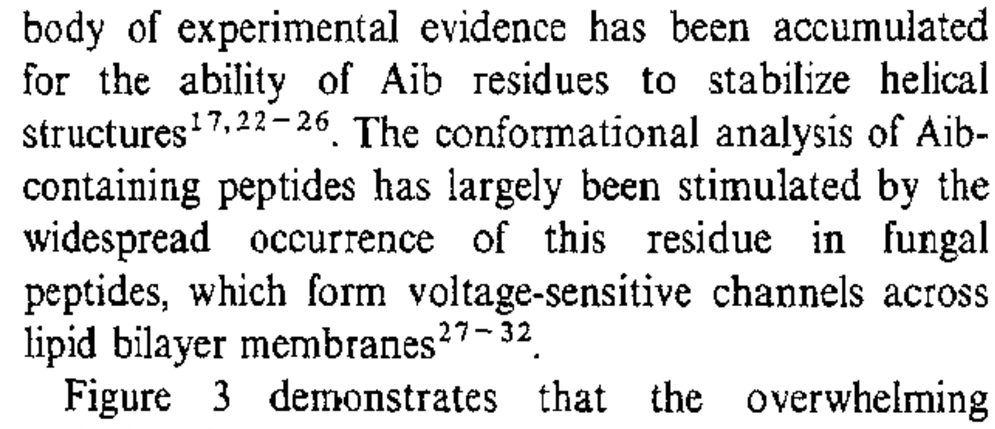

Figure 2. Structures of some α,α-dialkylated amino acids. Aib, α-aminoisobutyric acid; Deg, diethylglycine; Dpg, dipropylglycine; Ac_5c, 1-aminocyclopentane-1-carboxylic acid; Ac_6c, 1-aminocyclohexane-1-carboxylic acid.

Experimental section

The decapeptide was synthesized by conventional solution-phase procedures using a fragment condensation approach[46] and purified by reverse phase HPLC on a C_{18} column using a linear $MeOH/H_2O$ gradient. Crystals were grown by slow evaporation from isopropanol. The cell parameters of the polymorphs obtained from isopropanol and from methanol/water are compared in Table 1.

Table 1. Diffraction data for Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe.

	Polymorph from isopropanol/water	Polymorph from methanol/water (ref. 45)
Empirical formula	$C_{49}H_{88}N_{10}O_{13}\cdot C_3H_7OH$	$C_{49}H_{88}N_{10}O_{13}$
Crystallizing solvent	Isopropanol	Methanol/H_2O
Colour habit	Clear plate	Clear plate
Space group	$P2_1$	$P2_1$
Cell parameters (Å)	$a = 9.300(1)$	16.625(2)
	$b = 20.703(3)$	9.811(5)
	$c = 16.551(2)$	18.412(3)
	$\beta = 94.22(1)°$	99.79(1)°
Volume (Å^3)	3177.3	2959.4
Density, calc. (g cm^{-3})	1.134*	1.150
Formula weight	1025.2 + 60.1*	1025.2
Radiation, CuK_α (Å)	$\lambda = 1.54184$	1.54184
Temperature	Room	−71°C
Independent reflections	4141	5026
Observed reflections [$\lvert F_o\rvert > 3\sigma(F)$]	2956	4338
Final R indices (%) (observed data)	$R = 7.4$	$R = 5.7$
	$R_w = 7.7$	$R_w = 6.7$

*Assuming an equivalent of one molecule of C_3H_7OH per asymmetric unit.

X-ray diffraction data were measured on a dry crystal at room temperature with an automated four-circle diffractometer equipped with a graphite monochromator. The $\theta/2\theta$ scan mode was used with a $2.0° + 2\theta(\alpha_1-\alpha_2)$ scan, variable scan speeds depending on the intensity of the reflection and $2\theta_{max} = 110°$ (0.94 Å resolution). Three reflections, monitored after every 97 measurements, remained constant within 2% throughout the data collection. The structure was solved by a vector search procedure in the PATSEE computer program (ref. 47) contained in the SHELX84 package of programs (MicroVAX version of the SHELXTL system of programs, Siemens Instruments, Madison, WI, USA). The molecular model used in the vector search was the structure determined for the crystal grown from methanol[45]. Full-matrix, anisotropic least-squares refinement was performed on the C, N and O atoms in the peptide, after which hydrogen atoms were placed in idealized positions, with C–H = 0.96 Å, and allowed to ride with the C or N atom to which each was bonded for the final cycles of refinement. The thermal factor for the hydrogen atoms was fixed at $U_{iso} = 0.125$. Difference maps combined with further cycles of least-squares refinement were used to determine the contents of a large void in the cell present between helix molecules. The four strongest peaks in a difference map correspond to an isopropanol molecule at ~0.5 occupancy. Much weaker, although stable peaks (i.e. in the least-squares refinement) in the same area are difficult to interpret as another isopropanol molecule with a different orientation or perhaps, disordered water molecules. The final R factor for 2956 data (observed $> 3\sigma$) is 7.4%.

Fractional coordinates are listed in Table 2 and torsional angles are listed in Table 3.

All the NMR studies were carried out on a Bruker WH-270 FT-NMR spectrometer as described earlier[46]. CD spectra were recorded on a JASCO J500A spectropolarimeter using 0.1 mm pathlength cells. Signal averaging was performed and spectra are corrected for the solvent baseline obtained under the same conditions.

Results and discussion

Crystal structure

The helix. The decapeptide folds into a helix that is nearly identical in both crystal forms. A superposition of the molecular conformations from the two crystals, with a least-squares fit of the backbone atoms, is shown in Figure 4. The r.m.s. deviation in the positions of the pairs of 39 atoms from C′(0) to N(10) is only 0.22 Å. The values of the torsional angles ϕ differ by an average of less than 3° and those for ψ differ by an average of 5° in the molecules from the two crystals.

The backbone is completely α-helical with seven 5→1 type hydrogen bonds (Table 4 and Figure 5). In addition to the N(4)H. . . O(0) hydrogen bond where O(0) is part of the Boc end group, the N(3)H moiety also is directed toward O(0) for the probable formation of a 4→1 hydrogen bond in both crystal forms. The N . . . O bond lengths range from 2.89 to 3.22 Å. The C=O. . .N angle in the 5→1 type hydrogen bonds ranges from 143° to 168°.

The head-to-tail hydrogen bonding between the helices forms continuous columns of peptides in each crystal. Again, in both crystals the same bonds are formed, that is N(1)H. . .O(8) and N2(H). . .O(9).

The side chains. Although the helical backbones of the peptide are nearly identical in the two crystal forms, the conformations of the leucyl side chains are different. In Figure 4, it can be seen that in Leu(7), the side chain assumes the *g*, (*tg*) conformation in both crystals. For Leu(3), the side chain has the *t*, (tg^-) conformation in the present crystal and the *g*, (*tg*) conformation in the crystal without solvent molecules. For the Leu(9) side chain, the conformation is *g*, (*tg*) in the present crystal compared to an unusual χ^1 value of −90° ($\chi^2 = 175°, 50°$) in the other crystal. These variations in the conformations of the Leu side chains are undoubtedly related to the presence of nearest neighbours in the different modes of aggregation of the peptide columns in the two crystals (*vide infra*). Similarly, in the crystal of Boc–$(Leu_3–Aib)_2$–OBz, the Leu side chains adopted *t*, (tg^-) and *g*, (*tg*) conformations, as well as an unusual conformation due to packing considerations[48].

Table 2. Atomic coordinates ($\times 10^4$) and equivalent isotropic displacement coefficients ($Å^2 \times 10^3$).

	x	y	z	U(eq)*
C(1)	5656(21)	−2261(10)	−673(9)	78(7)
C(2)	5882(25)	−2523(11)	−1498(11)	135(11)
C(3)	6925(38)	−2377(18)	−139(20)	314(29)
C(4)	4481(43)	−2602(13)	−356(20)	278(27)
O	5321(13)	−1602(7)	−780(6)	84(5)
C′(0)	5318(15)	−1185(9)	−165(8)	59(6)
O(0)	4968(10)	−1309(7)	511(5)	66(4)
N(1)	5689(12)	−589(8)	−405(6)	55(4)
C^{α}(1)	6127(13)	−83(9)	180(7)	52(5)
C′(1)	4935(15)	35(8)	763(7)	50(5)
O(1)	5227(9)	65(7)	1507(4)	56(3)
C^{β}1(1)	7532(13)	−307(9)	667(8)	63(6)
C^{β}2(1)	6450(18)	538(9)	−262(9)	78(7)
N(2)	3629(12)	179(7)	422(6)	47(4)
C^{α}(2)	2496(14)	403(8)	912(7)	49(5)
C′(2)	2038(13)	−121(9)	1487(7)	48(5)
O(2)	1567(10)	14(7)	2143(5)	66(4)
C^{β}(2)	1229(16)	684(10)	392(9)	79(7)
N(3)	2087(11)	−729(8)	1210(6)	55(4)
C^{α}(3)	1678(14)	−1283(8)	1705(7)	52(5)
C′(3)	2804(15)	−1369(9)	2412(7)	52(5)
O(3)	2424(11)	−1504(7)	3093(5)	71(4)
C^{β}(3)	1484(20)	−1890(9)	1160(9)	77(7)
C^{γ}(3)	1057(19)	−2494(9)	1523(10)	79(7)
C^{δ}1(3)	931(23)	−2994(10)	810(12)	120(10)
C^{δ}2(3)	−321(18)	−2453(10)	1914(12)	101(8)
N(4)	4186(12)	−1311(8)	2267(6)	54(4)
C^{α}(4)	5315(15)	−1355(9)	2932(7)	64(6)
C′(4)	5165(13)	−793(9)	3543(8)	55(6)
O(4)	5361(10)	−926(7)	4272(5)	63(4)
C^{β}(4)	6797(14)	−1354(10)	2610(8)	74(6)
N(5)	4899(11)	−205	3258(6)	46(4)
C^{α}(5)	4753(14)	351(8)	3804(7)	50(5)
C′(5)	3617(14)	178(8)	4396(7)	48(5)
O(5)	3835(9)	253(7)	5128(5)	58(3)
C^{β}1(5)	4200(16)	919(9)	3277(8)	63(6)
C^{β}2(5)	6188(14)	520(9)	4270(8)	73(6)
N(6)	2351(10)	−42(7)	4068(6)	47(4)
C^{α}(6)	1123(13)	−173(9)	4536(7)	51(5)
C′(6)	1557(14)	−671(9)	5202(7)	52(5)
O(6)	1256(10)	−595(7)	5900(5)	69(4)
C^{β}1(6)	−76(14)	−468(9)	3948(8)	63(6)
C^{β}2(6)	567(16)	463(9)	4907(8)	69(6)
N(7)	2225(10)	−1205(7)	4960(5)	44(4)
C^{α}(7)	2478(14)	−1747(8)	5497(7)	51(5)
C′(7)	3758(17)	−1629(9)	6122(8)	59(6)
O(7)	3705(13)	−1871(8)	6793(6)	97(5)
C^{β}(7)	2643(17)	−2372(8)	5024(8)	64(6)
C^{γ}(7)	1275(20)	−2576(9)	4526(9)	76(7)
C^{δ}1(7)	1519(24)	−3163(10)	4075(11)	118(10)
C^{δ}2(7)	−33(27)	−2614(14)	4955(16)	157(13)
N(8)	4787(11)	−1248(8)	5914(5)	49(4)
C^{α}(8)	6014(14)	−1086(9)	6481(7)	55(5)
C′(8)	5555(12)	−635(9)	7159(7)	51(5)
O(8)	6213(9)	−641(7)	7825(5)	59(3)
C^{β}(8)	7212(14)	−784(11)	6054(8)	78(7)
N(9)	4503(10)	−210(8)	6952(6)	53(4)
C^{α}(9)	3947(14)	256(8)	7534(7)	51(5)
C′(9)	2863(13)	−40(9)	8064(8)	52(5)
O(9)	2602(10)	245(7)	8704(5)	70(4)
C^{β}(9)	3275(16)	856(9)	7108(8)	66(6)
C^{γ}(9)	4418(18)	1321(9)	6807(10)	77(7)
C^{δ}1(9)	3664(21)	1786(11)	6182(11)	104(8)
C^{δ}2(9)	5206(20)	1671(10)	7522(11)	107(9)
N(10)	2146(11)	−566(8)	7799(6)	57(4)
C^{α}(10)	923(12)	−822(9)	8180(7)	50(5)
C′(10)	−134(15)	−290(9)	8285(8)	55(5)
O(10)	−813(11)	−205(8)	8866(6)	85(5)
C^{β}1(10)	201(15)	−1342(9)	7627(8)	65(6)
C^{β}2(10)	1356(17)	−1148(11)	9008(8)	90(7)
O(11)	−343(10)	56(7)	7606(6)	69(4)
C(11)	−1334(20)	565(11)	7582(12)	110(9)
C(1S)	−1040(70)	1608(33)	2609(40)	177(22)
C(2S)	347(42)	1748(19)	2576(22)	119(12)
C(3S)	−1704(57)	1229(29)	2110(31)	173(19)
O(4S)	−1703(88)	1971(38)	2958(47)	158(30)
X(5S)	−1820(83)	1903(37)	2352(47)	168(26)
X(6S)	−2470(90)	1847(39)	2937(48)	39(20)
X(7S)	−345(106)	1041(51)	2805(58)	64(26)
X(8S)	−973(94)	840(41)	2442(52)	42(20)

*Equivalent isotropic U defined as one third of the trace of the orthogonalized U_{ij} tensor.

Table 3. Torsion angles[a,b], (deg).

Residue	ϕ	ψ	ω	χ^1	χ^2
Aib(1)	−57[c]	−56	−170		
Ala(2)	−66	−34	180		
Leu(3)	−69	−42	177	−179	179, 58
Ala(4)	−63	−43	180		
Aib(5)	−53	−51	−174		
Aib(6)	−58	−50	−170		
Leu(7)	−77	−29	177	−66	178, −52
Ala(8)	−71	−34	180		
Leu(9)	−80	−23	−169	−74	163, −72
Aib(10)	+51	+46[d]	178[e]		

[a]The torsion angles for rotation about bonds of the peptide backbone (ϕ, ψ and ω) and about bonds of the amino acid side chains (χ^1 and χ^2) described in ref. 51.
[b]E.s.d.'s 1.0°.
[c]C′(0), N(1), C^{α}(1), C′(1).
[d]N(10), C^{α}(10) C′(10), O(OMe).
[e]C^{α}(10), C′(10), O(OMe), C(OMe).

Torsion angles. It has been noted that values for the ϕ and ψ angles for α-helical peptides have a much wider range than previously anticipated[45]. Accordingly, a Ramachandran plot[1] has been constructed in Figure 6 for eight α-helical decapeptides consisting only of Aib(U), Ala(A), and Leu(L) residues using the values obtained from well-determined crystal structure analyses with resolutions of ~0.9 Å. These peptides are:

a) Boc–UALAUULALU–OMe (this paper; ref. 45)
b) Boc–UALALULALU–OMe (ref. 40)
c) Boc–UAUALALULU–OMe (ref. 45)
d) Boc–UALUALUALU–OMe (unpublished).

Peptides (a) and (b) each occur in two different crystal forms, while peptides (c) and (d) each occur with two independent molecules in the same crystal. The final Aib residue has been omitted from the plot, since in this series of peptides there usually is a helix reversal for the last residue.

The ϕ and ψ values for the Aib, Ala and Leu residues fall into three distinct clusters with a small amount of overlap. Overall the composite cluster in Figure 6 is elongated with smaller $|\psi|$ values associated with larger $|\phi|$ values. The inverse correlation between ϕ and ψ in regular helices has been noted by Barlow and Thornton

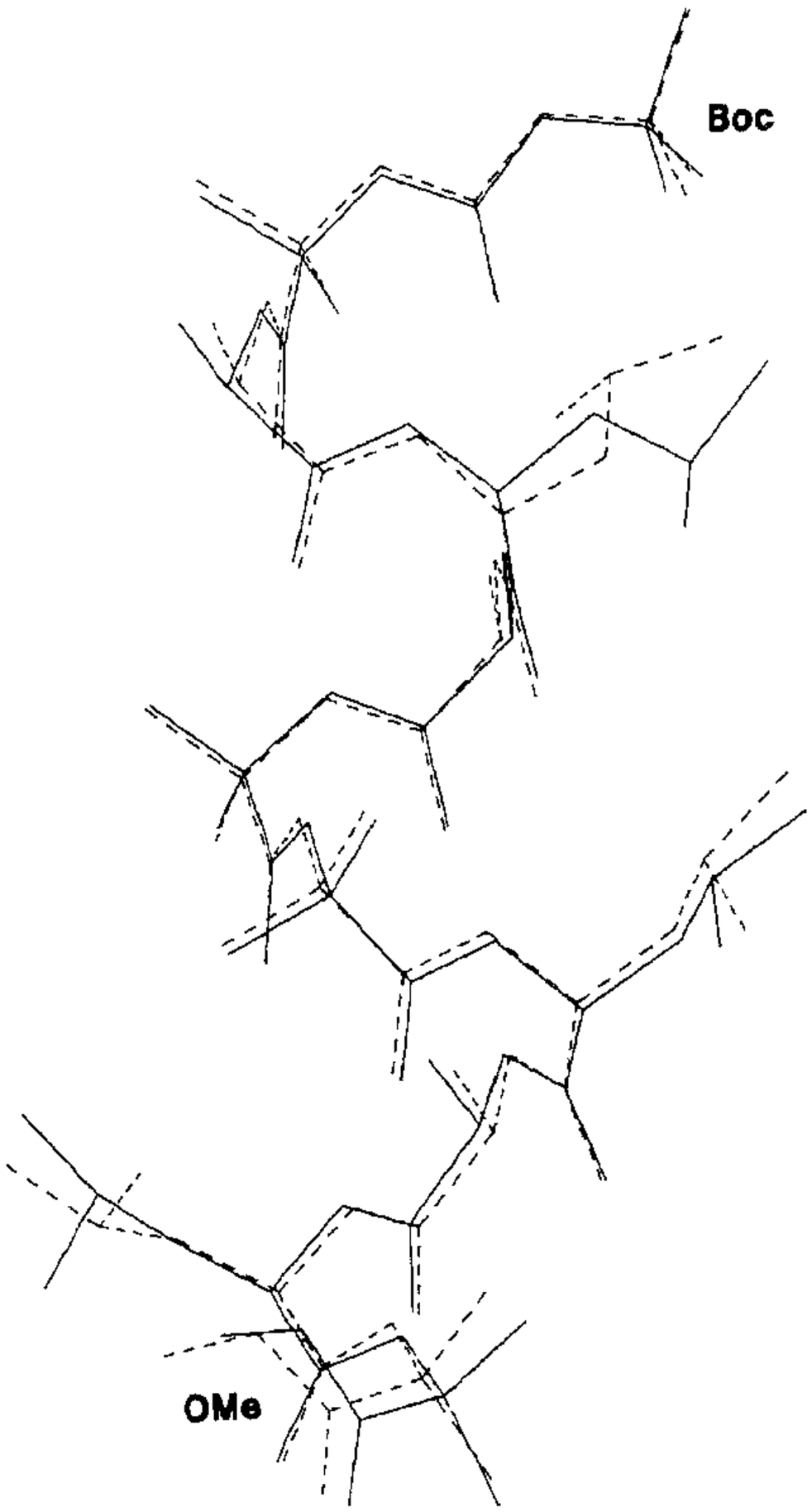

Figure 4. Superposition of decapeptide in crystal grown from isopropanol (solid line) with decapeptide in crystal grown from methanol (dotted line). The backbone atoms were fitted to each other by least squares.

Table 4. Hydrogen bonds.

Type	Donor	Acceptor	N. . .O(Å)	H. . .O[a](Å)	Angle (deg.) C=O. . .N
Head-to-tail	N(1)	O(8)[b]	3.006	2.11	141
	N(2)	O(9)[b]	2.935	2.00	137
3_{10}-Helix (4→1)	N(3)	O(0)	3.229	2.64	122
α-Helix (5→1)	N(4)	O(0)	3.048	2.10	168
	N(5)	O(1)	2.989	2.06	157
	N(6)	O(2)	3.217	2.29	143
	N(7)	O(3)	3.170	2.27	152
	N(8)	O(4)	2.887	1.98	161
	N(9)	O(5)	3.184	2.31	155
	N(10)	O(6)	3.191	2.57	151
Solvent-to-peptide	O(4S)[c]	O(7)	3.08		
	X(8S)[c]	O(2)	2.99		

[a] The H atoms were placed in idealized positions with the N–H distance equal to 0.96 Å.
[b] Symmetry equivalent $x, y, -1+z$ to coordinates listed in Table 2.
[c] At partial occupancy.

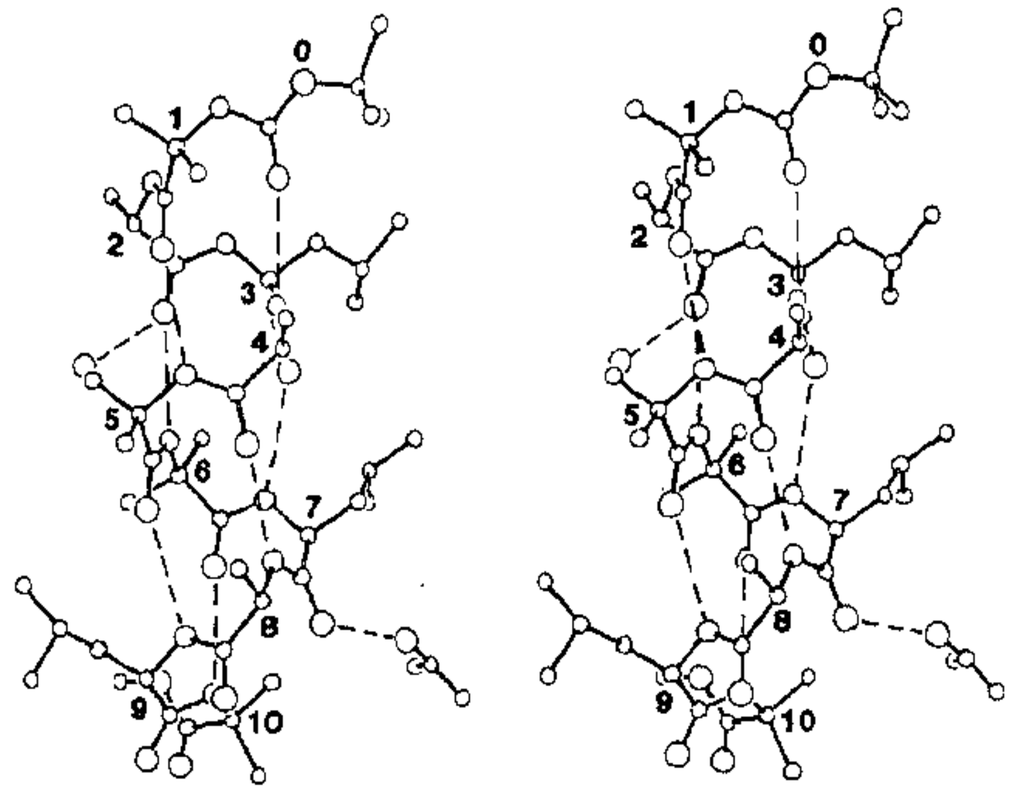

Figure 5. Stereodiagram of the α-helical Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe. A solvent molecule of isopropanol is shown on the right. An O atom from another possible isopropanol orientation is shown on the left. The C^{α} atoms are labelled 1 to 10. The figure 0 labels an oxygen atom in the Boc end group. Hydrogen bonds are indicated by dashed lines.

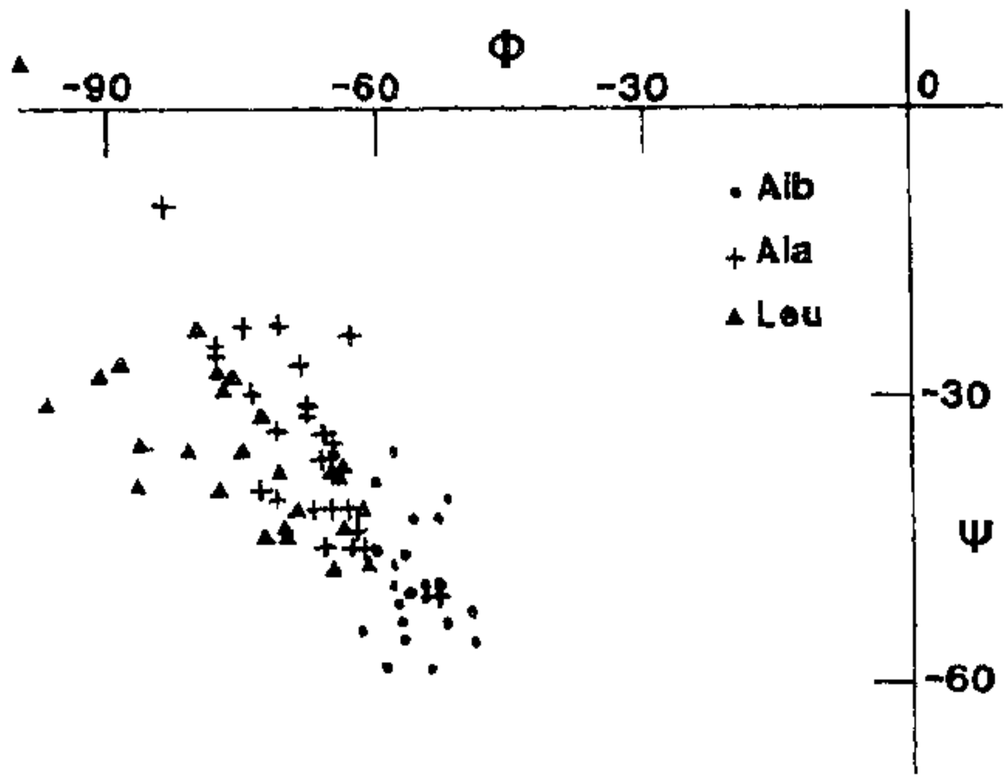

Figure 6. Plot of ϕ (N–C^{α}) and ψ (C^{α}–C′) values found in eight α-helical decapeptides with sequences containing only Aib, Ala and Leu residues. Each type of residue has been designated by an individual symbol: Aib ●, Ala + and Leu ▲.

in their study of helix geometry in proteins[49]. Each of the three residues, Aib, Ala and Leu, has an individual symbol in the plot. The median values of ϕ and ψ for Aib are near $-57°$ and $-50°$ and those for Ala are near $-67°$ and $-35°$. These values compare well with idealized values for α-helices[50] and theoretically computed values for Aib residues[18, 23]. The cluster for the Aib values is tighter than that for the Ala values. (There are ~24 data points in each cluster.) On the other hand, the cluster of ϕ and ψ values for the Leu residues spreads over at least twice the area than that for Ala. Excluding the outlier values of $\phi = -100°$ and $\psi = +5°$, the median is near $\phi = -75°$ and $\psi = -38°$. However, closer examination of the sequence around the Leu

residue shows that the ϕ and ψ values for Leu fall within the expected range, that is, superimposed on the cluster for Ala, when the Leu residue is between Ala residues or when the Leu–Aib sequence occurs within the body of the helix and that unusual ϕ and ψ values for Leu occur only when Aib immediately precedes Leu or when the Leu–Aib sequence is at the end of a helix. Nevertheless, these deviations from expected values are not large enough to disrupt an alpha-helix and normal 5→1 hydrogen bonds are formed.

Helix aggregation. The crystal grown from a methanol solution does not contain any cocrystallized solvent[45] while the one grown from isopropanol has the equivalent of one isopropanol molecule (disordered) cocrystallized with each peptide molecule. The common features in the packing of the two crystals are the continuous columns of helices formed by head-to-tail hydrogen bonding. These columns are nearly identical in conformation in the two crystals. They can be compared in Figures 7, *a* and *b* where they are projected end-on. The labels C and N indicate that the C or N termini are directed upwards. In each sheet of peptide columns, that is the repetition of columns in the vertical direction of both Figures 7, *a* and *b*, there is completely parallel association of the helices.

In both crystals, the helices in adjacent sheets run in antiparallel directions. But the 'N' sheets in Figure 7, *a* are turned by 180° compared to the 'N' sheets in Figure 7, *b*, which, of course, means that all intermolecular contacts between the sheets are different in the two crystals. Furthermore, the vertical sheets are spaced in the horizontal direction by a larger distance in Figure 7, *a* than Figure 7, *b*, 10.3 Å compared to 9.2 Å, in order to accommodate the isopropanol molecules.

The space between the helices is larger than necessary for an isopropanol molecule. As a consequence, the isopropanol molecule appears to be disordered among a number of different orientations. Only the site with occupancy of ~0.6 is shown in the figures. In this orientation the OH forms a hydrogen bond with O(7) of the helix. Another possible orientation for which there is some evidence among the minor peaks displayed in the difference map, is the rotation of the isopropanol so that the OH forms a hydrogen bond with O(2) of a helix on the other side of the solvent channel. Apart from the probable hydrogen bonding of the isopropanol, the C. . .C intermolecular distances between peptides and solvent are large, generally >4.2 Å. Figure 8 shows the antiparallel packing of the helices and the channels containing the isopropanol.

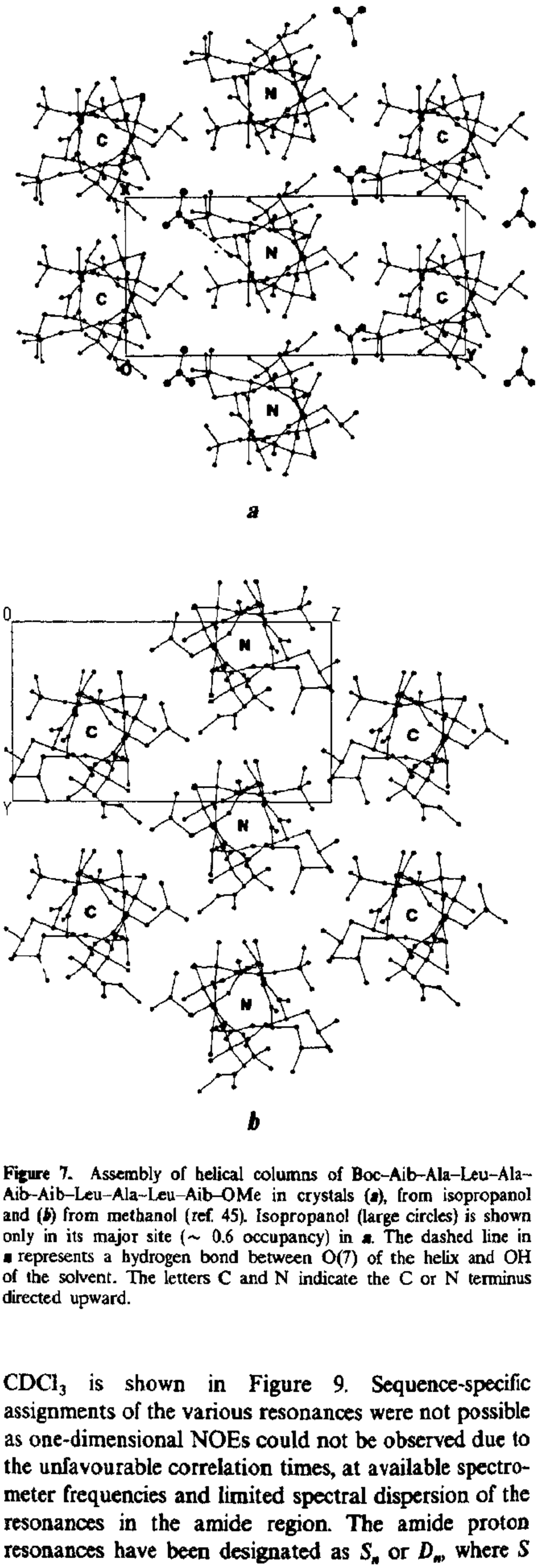

Figure 7. Assembly of helical columns of Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe in crystals (*a*), from isopropanol and (*b*) from methanol (ref. 45). Isopropanol (large circles) is shown only in its major site (~ 0.6 occupancy) in *a*. The dashed line in *a* represents a hydrogen bond between O(7) of the helix and OH of the solvent. The letters C and N indicate the C or N terminus directed upward.

Solution conformations

^{1}H NMR studies on I were carried out in $CDCl_3$ and $(CD_3)_2SO$. A two-dimensional COSY spectrum of I in $CDCl_3$ is shown in Figure 9. Sequence-specific assignments of the various resonances were not possible as one-dimensional NOEs could not be observed due to the unfavourable correlation times, at available spectrometer frequencies and limited spectral dispersion of the resonances in the amide region. The amide proton resonances have been designated as S_n or D_n, where S

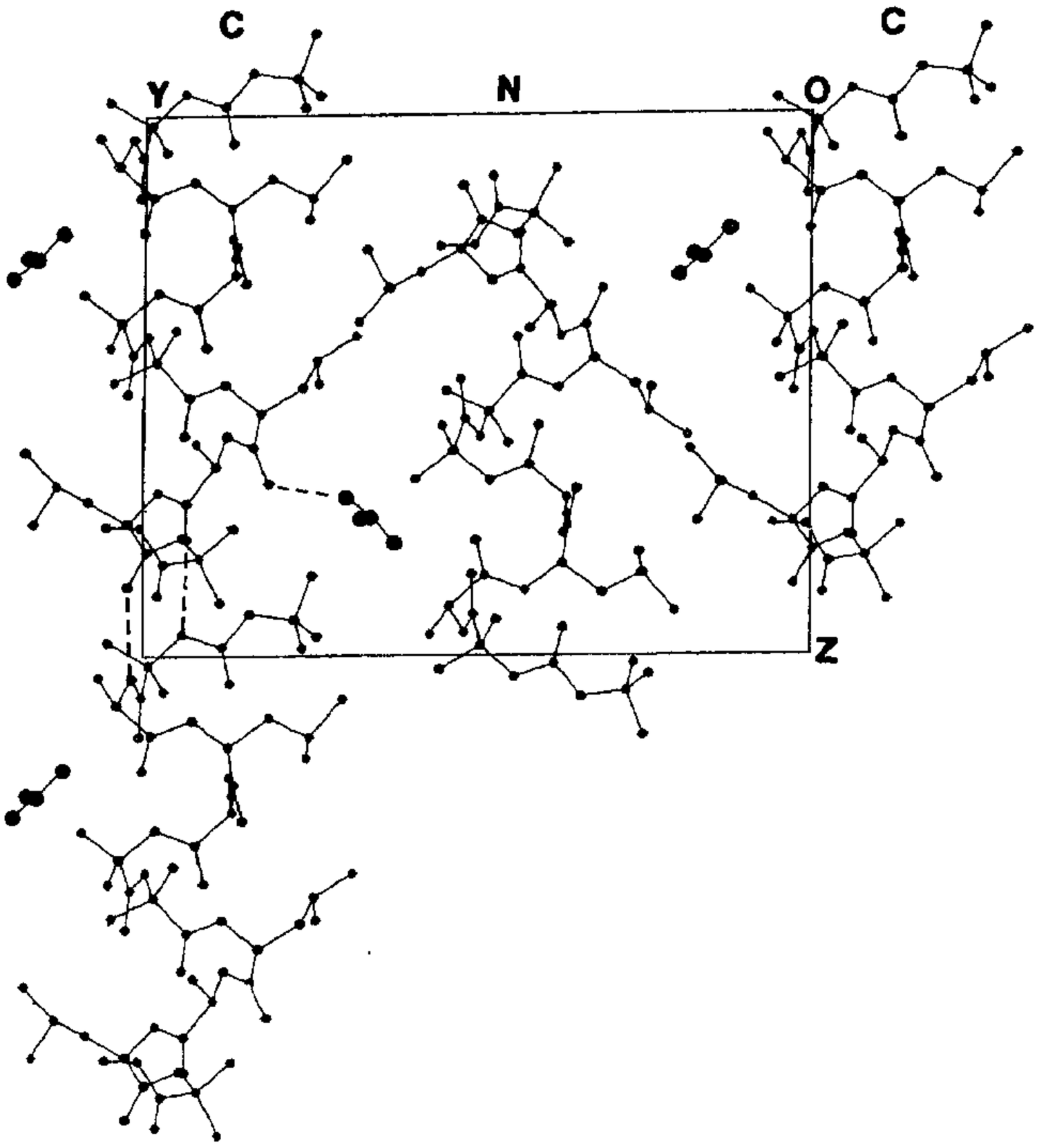

Figure 8. Antiparallel packing in Figure 7, *a* shown in a view perpendicular to the helices. Dashed lines represent head-to-tail hydrogen bonds, N(1)H. . .O(8) and N(2)H. . .O(9), and solvent to peptide hydrogen bond, OH. . .O(7). Isopropanol (large circles) is shown only in its major orientation.

and *D* refer to the multiplicity of the resonance and *n* refers to their order of appearance from lowfield in $(CD_3)_2SO$.

Solvent accessibility of the amide resonances has been probed using solvent perturbation (in $CDCl_3$–$(CD_3)_2SO$ mixtures) and temperature dependence (in $(CD_3)_2SO$), of NH resonance chemical shifts. A plot of the NH chemical shifts as a function of $CDCl_3$–$(CD_3)_2SO$ composition is shown in Figure 10. ^1H NMR parameters of the amide resonances in both the solvents are listed in Table 5.

Two of the NH resonances, D_1 and S_9, exhibit significant downfield chemical shifts with increasing concentrations of $(CD_3)_2SO$, suggesting the exposed nature of these groups. Two more singlet resonances, S_2 and S_4, exhibit marginal downfield chemical shifts which are more significant after $\sim 10\%$ of $(CD_3)_2SO$. The chemical shifts of the remaining resonances are markedly insensitive to solvent composition. In $(CD_3)_2SO$, D_1 and S_9 have large temperature coefficients (≥ 6.0 ppm $\times 10^{-3}$ K^{-1}), suggesting their solvent-exposed nature. S_4 has an intermediate coefficient (3.3 ppm $\times 10^{-3}$ K^{-1}) and the remaining resonances have low coefficients (≤ 3.0 ppm $\times 10^{-3}$ K^{-1}).

S_9 can be assigned to Aib(1)NH as urethane protons appear at higher fields compared to the remaining amide resonances, in $CDCl_3$ (ref. 46). A comparison of the solvent perturbation data with that for an isomeric sequence, Boc–Aib–(Ala–Leu–Aib)$_3$–OMe, for which sequence-specific assignments are available, suggests that D_1 is Ala(2)NH. The presence of eight solvent-shielded NH groups in $CDCl_3$ and seven in $(CD_3)_2SO$ is consistent with helical conformations in solution. In the latter solvent, one more Aib NH (S_4) is partially solvent exposed.

Further support for a helical conformation in solution is obtained from the coupling constants of the amide resonances. The $^3J_{HNC^{\alpha}H}$ values in both the solvents are in the range of 3 to 6.8 Hz, with the exception of one Leu NH (D_{10}, 8.6 Hz). In several of the

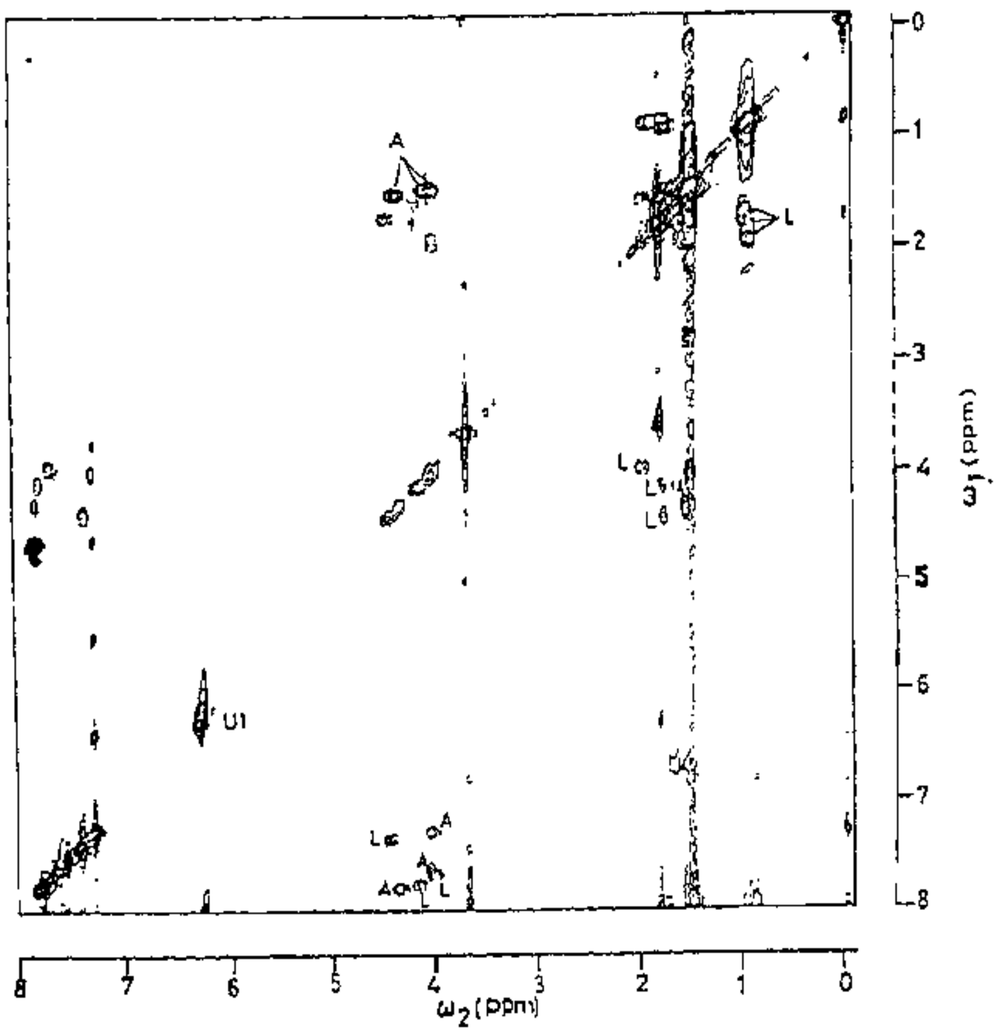

Figure 9. Contour plot of the 270 MHz COSY spectrum of I in $CDCl_3$.

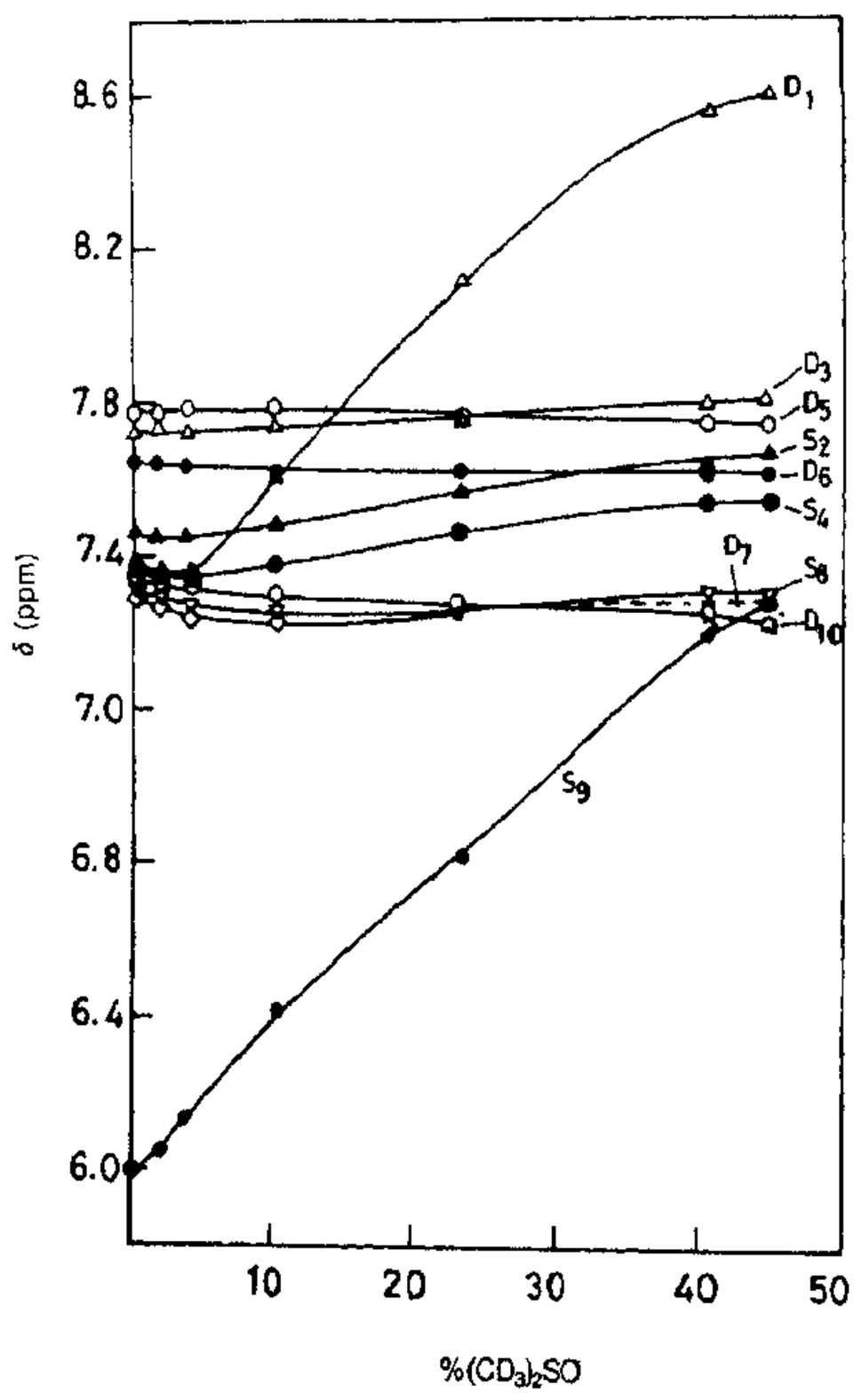

Figure 10. Solvent dependence of NH chemical shifts in I as a function of $(CD_3)_2SO$ concentration in $CDCl_3$–$(CD_3)_2SO$ mixtures.

Table 5. ^1H NMR parameters for amide resonances in peptide I[a].

Resonance	δ_{NH} (ppm) $CDCl_3$	$^3J_{HNC^\alpha H^\alpha}$ (Hz) $CDCl_3$	δ_{NH} (ppm) $(CD_3)_2SO$	$^3J_{HNC^\alpha H^\alpha}$ (Hz) $(CD_3)_2SO$	$d\delta/dT$ (ppm × 10^{-3} K^{-1}) $(CD_3)_2SO$
D_1	7.59	3.0[c]	8.34	6.1	6.0
S_2	7.53	—	7.93	—	3.0
D_3	7.76	6.5	7.82	*d*	2.0
S_4	7.40	—	7.82	—	3.3
D_5	7.80	6.2	7.70	6.4	1.3
D_6	7.67	5.1	7.61	6.1	1.3
D_7	7.32	6.8	7.61	*d*	0.8
S_8	7.38	—	7.43	—	7.0
S_9	6.22	—	7.43	—	7.0
D_{10}	7.39	6.5	7.18	8.6	1.3

[a]Peptide concentration ~11 mM.
[b]Errors in J values are ~±0.5 Hz.
[c]This coupling constant has been determined at a peptide concentration of 1.2 mM.
[d]J value could not be determined as the resonance is broad.

decapeptides of related sequences, which have been studied, it has been observed that one Leu NH has a large coupling constant and in the solid state, Leu(9)NH has a larger ϕ than the remaining residues. In the case of peptide I also, Leu(9)NH has a larger ϕ in both crystal forms (−91° or −80°). Coupling constants in the range 2 to 6 Hz are indicative of ϕ values in the helical region of conformational space[52].

The strength of the helix formed by I in the polar, strongly solvating $(CD_3)_2SO$ has been further probed by heating the sample to 373 K. The NH chemical shift vs temperature plots are linear over a wide range of temperature (273 to 373 K) and there is no change in the $C^\alpha H$ region, suggesting that the peptide retains a helical conformation, even at 373 K.

A CD spectrum of I in MeOH (Figure 11) gives band maxima at 222, 205 and 190 nm, with molar ellipticities of −4.1, −10.0 and 10.9×10^4 deg cm^2 $decimol^{-1}$, respectively, providing strong evidence for a helical conformation.

Conformational comparisons

The peptide Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib–OMe (I) adopts a largely α-helical conformation in two different crystal forms. In both structures, seven intramolecular 5→1 hydrogen bonds are observed. There is also one weak 4→1 interaction involving the Leu(3)NH and Boc CO groups in both cases. The Boc CO group simultaneously hydrogen bonds with both Leu(3)NH (4→1) and Ala(4)NH (5→1), a feature commonly observed at the N-termini of peptide helices. NMR studies, in solution, support a conformation with seven solvent-shielded NH groups and one partially exposed NH group. The NH groups of Aib(1) and Ala(2) are fully exposed. This is completely consistent with a predominantly helical conformation of the type

observed in crystals. Thus, the decapeptide I appears to be a stereochemically stable helix, with the environment having little influence on backbone conformation. Indeed, an Aib content of 30–40% appears to stabilize largely homogeneous, helical conformations as evidenced by studies on decapeptides with related sequences. Table 6 summarizes the crystallographic and NMR results on a set of similar peptides. The close identity of backbone conformations is also evident from CD studies in methanol (Figure 11), which clearly reveal exciton split π–π^* bands at 205 nm and 190 nm and an n–π^* band at 222 nm. The observed ellipticities are characteristic of helical peptides of this length.

Implications for design

The growing body of evidence on the helical conformations of Aib-containing peptides leads to some general conclusions:

(i) Well-defined helical conformations ranging from 2

Table 6. Structural features of peptide helices.

	Solution conformation				Solid state conformation		
	NMR		CD				
	Number of solvent shielded amide groups[b]				Hydrogen bonds		
Peptide[a]	$CDCl_3$	$(CD_3)_2SO$[c]	$[\theta]/[\theta]$[d]	Secondary structure	Type	Number	Packing[e]
Boc–UALAUULALU–OMe (I)	8	7 and 1	0.41 −1.09 −2.66	α-Helix[f]	4→1 5→1	1 7	Antiparallel
				α-Helix[g]	4→1 5→1	1 7	Antiparallel[45]
Boc–UAUALALULU–OMe	8	6 and 2	0.29 −0.91 −3.11	α-Helix[g,h]	5→1	7	Skewed[45]
				α-Helix	4→1 5→1	2 6	
Boc–U(ALU)$_3$–OMe	8	8	0.24 −0.85 −3.56	Mixed[g] 3_{10}/α-helix	4→1 5→1 To inserted water	3 4 1	Antiparallel[34]
Boc–UALALULALU–OMe	8	7 and 1	0.41 −1.22 −2.96	α-Helix[g]	4→1 5→1	1 6	Antiparallel[40]
				α-Helix[f]	4→1 5→1	2 6	Parallel[40]
Boc–(MVULA)$_2$–OMe	8	5 and 3	0.47 −1.48 −3.18	Helix[g,h,i]			

[a]Single letter code: U, Aib; A, Ala; L, Leu; V, Val; M, Met.
Boc, *t*-butyloxycarbonyl; OMe, methyl ester.
[b]Number of solvent-shielded amide groups determined by solvent perturbation of amide resonances in $CDCl_3$–$(CD_3)_2SO$ mixtures (for $CDCl_3$) and temperature dependence of amide resonances [for $(CD_3)_2SO$].
[c]In $(CD_3)_2SO$, temperature coefficients ≤ 3 ppm $\times 10^{-3}$ K^{-1} are considered as characteristic of solvent-shielded and those in the range of 3 to 4.5 ppm $\times 10^{-3}$ K^{-1} as characteristic of partially exposed amide groups. The first number refers to shielded groups and second, to partially exposed groups.
[d]CD spectra have been recorded in methanol. The three transitions characteristic of helical conformations, n–π^* (~220 nm) and π–π^*// (~202 nm) and π–$\pi^*\perp$ (~190 nm) are seen in each case. The numbers given in the table refer to the ratios $[\theta]_{n-\pi^*}/[\theta]_{\pi-\pi^*\parallel}$, $[\theta]_{n-\pi^*\perp}/[\theta]_{\pi-\pi^*\perp}$ and $[\theta]_{\pi-\pi^*\perp}/[\theta]_{n-\pi^*}$, respectively.
[e]All the peptides have crystallized in the space group $P2_1$.
[f]Crystals obtained from isopropanol/water.
[g]Crystals obtained from methanol/water.
[h]Two molecules in the asymmetric unit.
[i]Structure not yet determined.

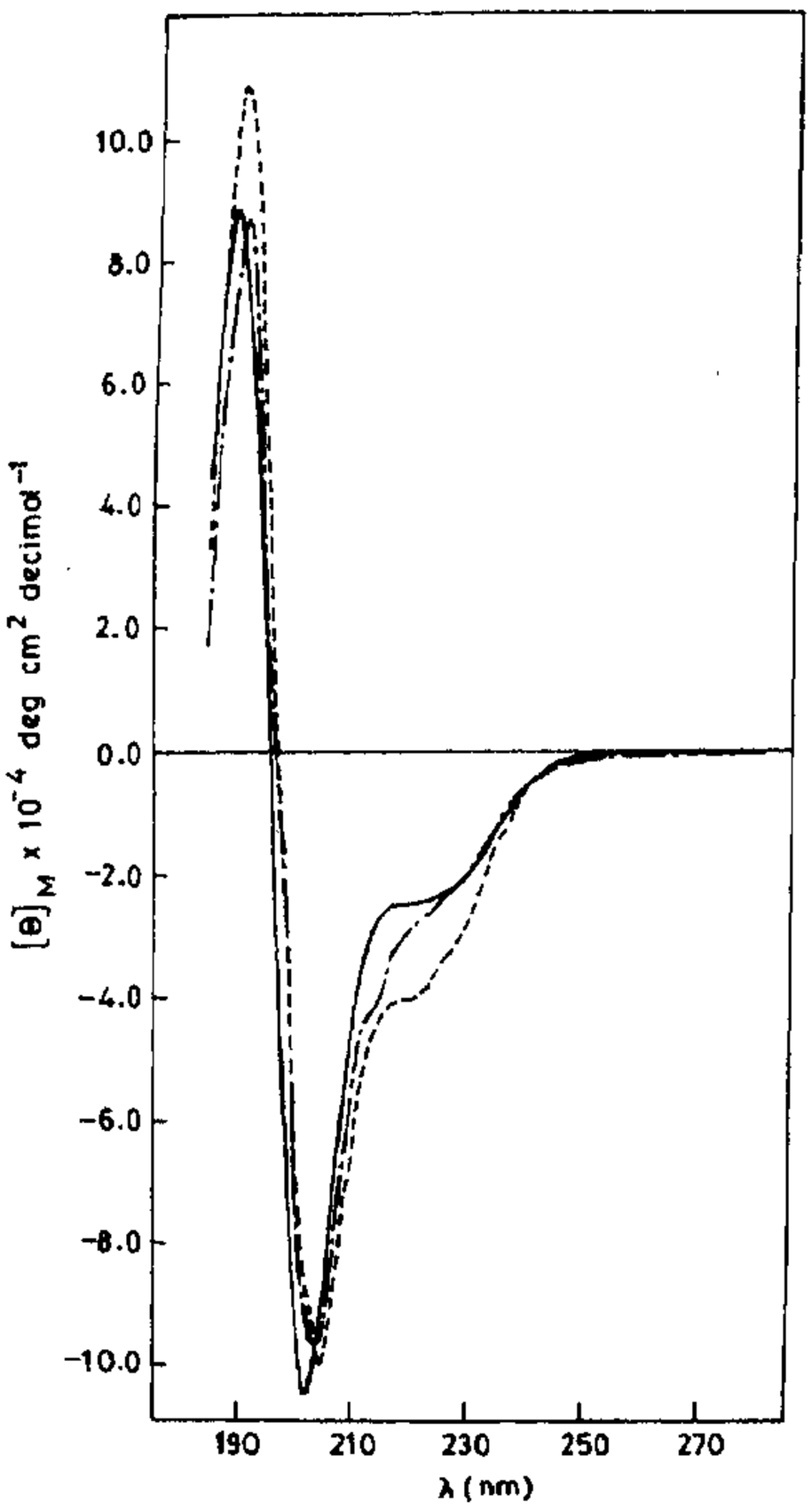

Figure 11. Circular dichroism spectra of I and two isomeric compounds, in methanol. Peptide concentration, 1 mg/ml. ---, Boc–Aib–Ala–Leu–Ala–Aib–Aib–Leu–Ala–Leu–Aib-OMe (I); ——, Boc–Aib–Ala–Leu–Aib–Ala–Leu–Aib–Ala–Leu–Aib-OMe; –·–, Boc–Aib–Ala–Aib–Ala–Leu–Ala–Leu–Aib–Leu–Aib–OMe.

to 4 turns of an α-helix can be readily constructed and characterized in crystals and in solution. While the precise determinants of 3_{10} or α-helical conformations continue to be debated[43], the distinctions between the two helix types are subtle enough to be ignored in the initial phases of a synthetic design project.

(ii) The synthetic peptide helices are generally stable in apolar organic solvents like chloroform. In more strongly solvating media like dimethylsulphoxide, helix stability appears to be a function of Aib content. Peptides with ⩾ 30% Aib yield stable helices, while a lower Aib content results in helix destabilization in polar solvents. This is presumably, a consequence of the formation of solvent–solute hydrogen bonds resulting in opening of the helical backbone. Such effects have been noted in the sequence Boc–(Val–Ala–Leu–Aib)$_2$–OMe and Boc–Aib–(Val–Ala–Leu–Aib)$_2$–OMe.

(iii) Sequence effects appear to be of limited importance in the cases examined. The precise positioning of Aib does not appear to be a major determinant of helix formation. Several contiguous non-Aib residues can be accommodated into helices. The largest non-Aib stretch characterized so far is six residues in the solid state conformation of the 15-residue peptide Boc–Val–Ala–Leu–Aib–Val–Ala–Leu–Val–Ala–Leu–Aib–Val–Ala–Leu–Aib–OMe (ref. 44).

Peptide helices of lengths comparable to those generally found in globular proteins can thus be reproducibly constructed from Aib-containing sequences. Initial attempts to design an α,α-motif have led to the synthesis of the 18-residue peptide Boc–Aib–Val–Ala–Leu–Aib–Val–Ala–Leu–Gly–Pro–Val–Ala–Leu–Aib–Val–Ala–Leu–Aib–OMe. This sequence incorporates a central Gly–Pro segment as a potential hinge. Preliminary characterization of molecular dimensions by determination of the space group and unit cell dimensions of a crystalline form of the peptide indicates formation of a continuous helix. This conclusion is also supported by NMR analysis of the hydrogen bonding pattern and estimation of ϕ, ψ torsion angles at Gly from geminal and vicinal coupling constants (unpublished results). Further work in this area will focus on the assembly of helices into supersecondary structure motifs by appropriate choice of linker peptide sequences. The availability of sterochemically constrained amino acid residues with conformational properties distinct from Aib will greatly facilitate a rational approach to controlling orientation of secondary structure elements in a modular approach to synthetic protein design. Expanding the repertoire of amino acids with strong preferences for specific regions of the Ramachandran map appears to be a desirable goal.

1. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. *J. Mol. Biol.*, 1963, **7**, 95.
2. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.*, 1968, **23**, 284.
3. Richardson, J., *Adv. Protein Chem.*, 1981, **34**, 164.
4. Regan, L. and DeGrado, W. F., *Science*, 1988, **241**, 976.
5. DeGrado, W. F., Wasserman, Z. R. and Lear, J. D., *Science*, 1989, **243**, 622.
6. DeGrado, W. F., *Adv. Protein Chem.*, 1988, **39**, 51.
7. Richardson, J. S. and Richardson, D. C., *Protein Engineering* (eds. Oxender, D. L. and Fox, C. F.), Alan R. Liss, New York, 1987, pp. 149–163.
8. Mutter, M. and Vuilleumier, S., *Angew. Chem. Int. Ed. Engl.*, 1989, **28**, 535.
9. Moser, R., Frey, S., Munger, K., Hehlgans, T., Klauser, S., Langen, H., Winnacker, E., Metz, R. and Gutte, B., *Protein Eng.*, 1987, **1**, 339.
10. Urry, D. W., *Research and Development*, 1988, 72.
11. Kaumaya, P. T. P., Berndt, K. D., Heidorn, D. B., Trewhella, J., Kezdy, F. J. and Goldberg, E., *Biochemistry*, 1990, **29**, 13.
12. Fasman, G. D., *Trends Biochem. Sci.*, 1989, **14**, 295.

13. Robson, B. and Garnier, J., *Introduction to Proteins and Protein Engineering*, Elsevier, Amsterdam, 1988.
14. Balaram, P., *Proc. Indian Acad. Sci. (Chem. Sci.)*, 1984, **93**, 703.
15. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Biochemistry*, 1989, **28**, 6696.
16. Leszczynski, J. F. and Rose, G. D., *Science*, 1986, **234**, 849.
17. Toniolo, C. and Benedetti, E., *ISI Atlas of Science: Biochemistry*, 1988, **1**, 225–230.
18. Marshall, G. R. and Bosshard, H. E., *Circ. Res.*, 1972, **30/31**, (suppl. II), 143.
19. Shamala, N., Nagaraj, R. and Balaram, P., *Biochem. Biophys. Res. Commun.*, 1977, **79**, 292.
20. Nagaraj, R., Shamala, N. and Balaram, P., *J. Am. Chem. Soc.*, 1979, **101**, 16.
21. Shamala, N., Nagaraj, R. and Balaram, P., *J. Chem. Soc. Chem. Commun.*, 1978, 996.
22. Toniolo, C., Bonora, G. M., Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V. and Pedone, C., *Biopolymers*, 1983, **22**, 205.
23. Prasad, B. V. V. and Balaram, P., *CRC Crit. Rev. Biochem.*, 1984, **16**, 307.
24. Bosch, R., Jung, G., Schmitt, H. and Winter, W., *Biopolymers*, 1985, **24**, 961.
25. Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Tomolo, C., Bonora, G. M., Formaggio, F. and Crisma, M., *J. Biomol. Struc. Dyn.*, 1988, **4**, 803.
26. Uma, K. and Balaram, P., *Indian J. Chem.*, 1989, **28B**, 705.
27. Nagaraj, R. and Balaram, P., *Acc. Chem. Res.*, 1981, **14**, 356.
28. Mathew, M. K. and Balaram, P., *Mol. Cell. Biochem.*, 1983, **50**, 47.
29. Menestrina, G., Klaus, V.-P., Jung, G. and Boheim, G., *J. Membr. Biol.*, 1986, **93**, 111.
30. Hall, J. E., Vodyanov, I., Balasubramanian, T. E. and Marshall, G. R., *Biophys. J.*, 1984, **45**, 233.
31. Karle, I. L., Flippen-Anderson, J. L., Sukumar, M. and Balaram, P., *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 5087.
32. Krishna, K., Sukumar, M. and Balaram, P., *Pure Appl. Chem.*, 1990, **62**, 1417.
33. Karle, I. L., Sukumar, M. and Balaram, P., *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 9284.
34. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 299.
35. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Int. J. Pept. Protein Res.*, 1988, **32**, 536.
36. Karle, I. L., Flippen-Anderson, J. L., Uma, K., Balaram, H. and Balaram, P., *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 765.
37. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Proteins: Structure, Function and Genetics*, 1990, **7**, 62.
38. Karle, I. L., Flippen-Anderson, J. L., Sukumar, M. and Balaram, P., *Int. J. Pept. Protein Res.*, 1988, **31**, 567.
39. Karle, I. L., Flippen-Anderson, J. L., Sukumar, M. and Balaram, P., *Int. J. Pept. Protein Res.*, 1990, **35**, 518.
40. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Biopolymers*, 1990, **29**, 1835.
41. Bosch, R., Jung, G., Schmitt, H. and Winter, W., *Biopolymers*, 1985, **24**, 979.
42. Francis, A. K., Iqbal, M., Balaram, P. and Vijayan, M., *FEBS Lett.*, 1983, **155**, 230.
43. Marshall, G. R., Hodgkin, E. E., Langs, D. A., Smith, G. D., Zabrocki, J. and Leplawy, M. T., *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 487.
44. Karle, I. L., Flippen-Anderson, J. L., Uma, K., Sukumar, M. and Balaram, P., *J. Am. Chem. Soc.*, (in press).
45. Karle, I. L., Flippen-Anderson, J. L., Uma, K. and Balaram, P., *Biopolymers*, (in press).
46. Balaram, H., Sukumar, M. and Balaram, P., *Biopolymers*, 1986, **25**, 2209.
47. Egert, E. and Sheldrick, G. M., *Acta Crystallogr.*, 1985, **A41**, 262.
48. Okuyama, K., Tanaka, N., Doi, M. and Narita, M., *Bull. Chem. Soc. Jpn.*, 1988, **61**, 3115.
49. Barlow, D. J. and Thornton, J. M., *J. Mol. Biol.*, 1988, **201**, 601.
50. Chothia, C., *Annu. Rev. Biochem.*, 1984, **53**, 537.
51. IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, 1970, **9**, 3471.
52. Pardi, A., Billeter, M. and Wüthrich, K., *J. Mol. Biol.*, 1984, **180**, 741.

ACKNOWLEDGEMENTS. This research was supported in part by National Institute of Health Grant GM30902, by the Office of Naval Research, and by a grant from the Department of Science and Technology, India. K.U. is the recipient of a fellowship from the Council of Scientific and Industrial Research, New Delhi.