Helix Geometry in Proteins

D. J. Barlow† and J. M. Thornton

Department of Crystallography
Birkbeck College London
Malet Street
London WC1E 7HX, U.K.

(Received 12 June 1987, and in revised form 26 January 1988)

In this report we describe a general survey of all helices found in 57 of the known protein crystal structures, together with a detailed analysis of 48 α-helices found in 16 of the structures that are determined to high resolution.

The survey of all helices reveals a total of 291 α-helices, 71 3_{10}-helices and no examples of π-helices. The conformations of the observed helices are significantly different from the "ideal" linear structures. The mean ϕ, ψ angles for the α- and 3_{10}-helices found in proteins are, respectively, (−62°, −41°) and (−71°, −18°).

A computer program, HBEND, is used to characterize and to quantify the different types of helix distortion. α-Helices are classified as regular or irregular, linear, curved or kinked. Of the 48 α-helices analysed, only 15% are considered to be linear; 17% are kinked, and 58% are curved.

The curvature of helices is caused by differences in the peptide hydrogen bonding on opposite faces of the helix, reflecting carbonyl-solvent/side-chain interactions for the exposed residues, and packing constraints for residues involved in the hydrophobic core. Kinked helices arise either as a result of included proline residues, or because of conflicting requirements for the optimal packing of the helix side-chains.

In α-helices where there are kinks caused by proline residues, we show that the angle of kink is relatively constant (~26°), and that there is minimal disruption of the helix hydrogen bonding. The proline residues responsible for the kinks are highly conserved, suggesting that these distortions may be structurally/functionally important.

1. Introduction

Helices are the most common secondary structures found in globular proteins, accounting for just over one-third of all residues (see below). Knowledge of the details of their structure, and, in particular, knowledge about the commonly observed distortions, is therefore essential to our understanding of protein structure, function and folding.

Previous analyses have concentrated on the packing arrangements of α-helices (Chothia *et al.*, 1977; Richmond & Richards, 1978), and the tendency has been to regard these as regular, linear structures, which approximate to rigid cylindrical rods. Although individual authors have commented on the differences between globular protein-helices and the "standard" structures (Artymiuk & Blake, 1981; Takano & Dickerson, 1981), on the irregularity of their termini (Bolin *et al.*, 1982), and on the distortions caused by proline (Love *et al.*, 1971), serine and threonine residues (Bolin *et al.*, 1982), it is surprising that no systematic analysis of the regularity has yet been performed.

In this report we describe a general survey of all helices found in 57 of the known protein crystal structures, together with a more detailed analysis of the 48 α-helices found in 16 of the structures that are refined to high resolution.

2. General Analysis

(a) α-helices

An analysis of 291 helices defined according to Kabsch & Sander (1983) (using co-ordinates available from the Cambridge Protein Databank (Bernstein *et al.*, 1977), gave the mean helix length as approximately ten residues (see Fig. 1(a)). This corresponds to a helix of approximately three turns, with an end-to-end distance of ~15 Å (1 Å = 0·1 nm) (cf. Kabsch & Sander, 1983).

† Present address and address for correspondence: Departments of Pharmacy & Pharmacology, King's College London (KQC), Manresa Road, London SW3 6LX, UK.

0022-2836/88/110601-19 $03.00/0

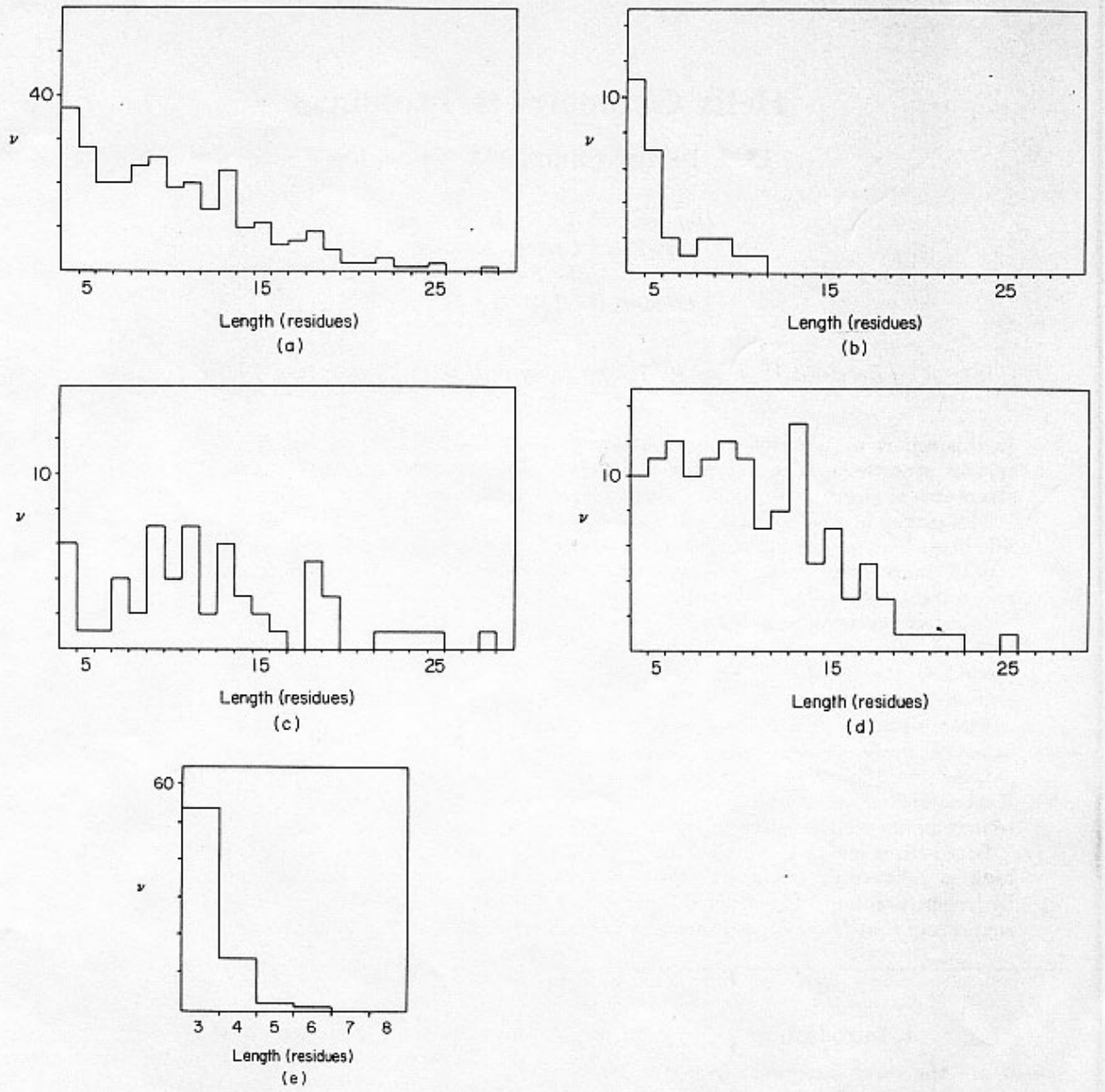

Figure 1. The distribution of observed helix lengths. ν represents the number of helices with a given number of residues. (a) α-Helices total sample (291 helices) (mean length = 10 res.). (b) α-Helices, all-β proteins (mean length = 5·7 res.). (c) α-Helices, all-α proteins (mean length = 12·2 res.). (d) α-Helices, α/β proteins (mean length = 10·2 res.). (e) 3_{10}-Helices, total sample (71 helices) (mean length = 3·3 res.).

In proteins composed principally of β-sheet (all-β proteins) the distribution of helix lengths is very narrow (see Fig. 1(b)), with 75% of the helices having <2 turns, and no helices longer than 11 residues. This contrasts with the distribution for all-α type proteins (Fig. 1(c)), where 20% of the helices have more than 17 residues, and only 12% are two turns or shorter. All lengths of helix are observed in the α/β proteins (Fig. 1(d)).

The averaged conformational parameters for the 291 helices are compared to various "standard" structures, in Table 1. The differences observed in the main-chain torsion angles result in a greater outward tilt of the carbonyl groups from the helix axis (see Fig. 2(a)), an effect that is particularly pronounced for the residues exposed to solvent (see below and Perutz *et al.*, 1965; Watson, 1969).

(b) 3_{10} *and* π *helices*

Of the 11,096 residues in the dataset, 376 (3·4%) are involved in 71 3_{10} helices. This compares with 3492 residues (32% of the total) that are involved in 291 α-helices. As would be expected on the basis of their relative stabilities therefore (Donoghue, 1953; De Santis *et al.*, 1965), 3_{10} helices are much less common than α-helices, but they are by no

Table 1

A comparison between various "standard" helices and those found in globular proteins

Helix	ϕ	ψ	p	n	r
A. *α-helices*					
Pauling *et al.* (1951)	−48°	−57°	5·5	3·65	2·3
Perutz (1951)	−67°	−44°	5·2	3·67	2·4
Arnott & Wonacott (1966)	−57°	−47°	5·5	3·59	2·3
Mean values for α-helices in globular proteins	−62°	−41°	5·4	3·54	2·3
B. *3_{10}-helices*					
Pauling *et al.* (1951)	−74°	−4°	6·0	3·0	1·8
Perutz (1951)	−49°	−26°	5·8	3·0	1·9
Poly-aib (Bavoso *et al.* 1986)	−54°	−28°	5·8	3·1	1·9
Mean values for 3_{10} helices in globular proteins	−71°	−18°	5·8	3·2	2·0

r is the radius of a helix (in Å), p is the pitch (in Å), and n is the number of residues per turn. ϕ and ψ are the polypeptide main-chain torsion angles.

means rare. Only a minority (24%) occur as an N- or C-terminal extension to an α-helix.

The distribution of 3_{10} helix lengths (Fig. 1(e)) shows that the majority are very short, with 96% having ≤4 residues. Rather surprisingly, however, 3_{10} helices are found to be significantly more common in the all-β proteins, than in the α/β proteins: the ratios of the observed : expected numbers of residues (calculated from the 3·4% occurrence) are 114 : 76 for all-β proteins, and 36 : 87 for α/β proteins. (A χ^2 analysis of the data, shows that the probability of these differences arising by chance, is <0·5%.) This suggests that the 3_{10} conformation is favourable for the connections between two β-strands, but unfavourable for the connections between α-helices and β-strands.

The 3_{10} helices that are observed in proteins show a wide spread in their residue ϕ, ψ angles (Fig. 2), and are generally irregular. They differ from the standard 3_{10} helices (Pauling *et al.*, 1951; Perutz, 1951) in that they have a larger radius and a smaller pitch (see Table 1), and are more closely related to the helical poly-aib structure (Bavoso *et al.*, 1986). Since the observed structures have 3·2 residues per turn, they are intermediate between an "ideal" 3_{10} helix (for which the number of residues per turn, n, is 3), and an α-helix (for which $n \sim 3{\cdot}6$).

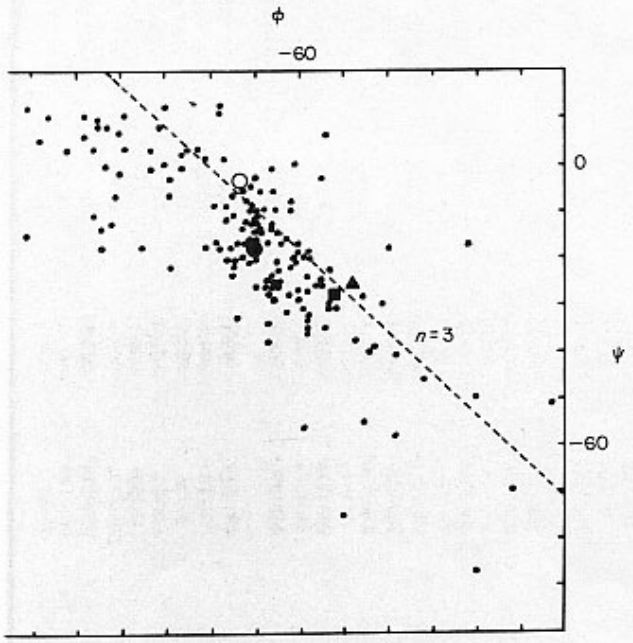

Figure 2. A Ramachandran plot showing the scatter of ϕ, ψ angles for residues in 3_{10}-helices. Symbols (○) and (▲) indicate the ϕ, ψ angles, respectively, for the "ideal 3_{10} helices of Pauling *et al.* (1951) and Perutz (1951); (■) the mean ϕ, ψ angles of the poly-aib helix (Bavoso *et al.*, 1986); (●) the mean ϕ, ψ angles for protein 3_{10}-helices. ϕ, ψ angles that generate helices with $n=3$ (i.e. 3 residues/turn) are indicated by a broken line.

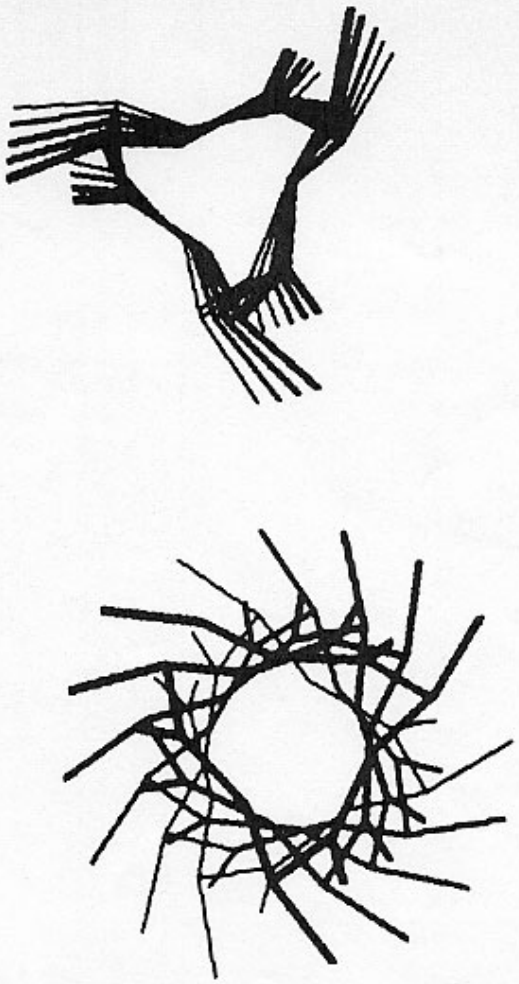

Figure 3. A comparison of the side-chain staggering for an "ideal" 3_{10}-helix (with ϕ, $\psi = -74°$, $-4°$, top diagram (Pauling *et al.*, 1951) and a 3_{10}-helix constructed with the mean ϕ, ψ angles of all protein 3_{10}-helices (ϕ, $\psi = -71°$, $-18°$, bottom diagram). The 2 helices are viewed down the helix axis, with the C^{α}–C^{β} bonds projecting radially outwards.

Table 2

The classification of α-helices according to their regularity and curvature

Protein	Helix	⟨φ⟩	⟨ψ⟩	⟨ω⟩	R_c(Å)	R_l(Å)	Radius (Å)	Classification	Comments
Lysozyme	5–15†	−65(6)	−38(9)	179(5)	0·25	0·25	26	Irregular	
	24–35*	−64(8)	−40(9)	−178(3)	0·11	0·28	58	Curved	
	89–100*	−62(7)	−42(3)	179(3)	0·08	0·32	57	Curved	
Ribonuclease	3–12*	−65(7)	−40(9)	179(4)	0·09	0·23	29	Curved	Slight C-terminal kink
	24–33	−68(9)	−41(9)	−179(3)	0·08	0·25	34	Curved	
Chymoyrypsin	235–244*†	−70(10)	−38(9)	−179(3)	0·05	0·13	62	Curved	
Carboxypeptidase	14–28*	−62(4)	−43(3)	180(4)	0·12	0·27	62	Curved	
	72–89*	−63(7)	−41(8)	179(5)	0·16	0·63	34	Kinked	Kink at Thr75
	173–187	−63(10)	−44(11)	180(4)	0·30	0·19	71	Irregular	Slight kink at Val176
	215–231*	−62(8)	−44(6)	180(6)	0·09	0·19	111	Linear	
	286–306*†	−61(6)	−43(7)	180(6)	0·22	0·69	45	Kinked	Kink at C terminus 286–305 = irregular
Thermolysin	67–88*	−64(9)	−44(7)	180(2)	0·21	0·43	89	Kinked	Kinked at N and C terminus
	136–151*	−63(5)	−40(9)	179(1)	0·15	0·46	35	Kinked	Kink at Val139
	159–180*†	−64(6)	−42(7)	179(1)	0·24	0·29	127	Kinked	Kink at Glu166
	234–246*†	−62(4)	−45(5)	179(1)	0·10	0·16	78	Curved	Slight C-terminal kink
	260–273*†	−64(7)	−44(8)	180(2)	0·17	0·22	71	Kinked	Kink at C terminus 260–272 = curved
	281–296*†	−66(8)	−42(6)	179(1)	0·14	0·17	100	Linear	Slight kink at C terminus
	300–313	−62(4)	−42(6)	179(1)	0·08	0·20	69	Curved	
Myoglobin	3–18	−64(9)	−39(10)	179(4)	0·14	0·34	49	Curved	
	20–36	−65(7)	−43(7)	180(4)	0·76	0·19	41	Irregular	
	58–77	−61(6)	−43(7)	180(5)	0·11	0·28	85	Curved	
	82–95*	−69(7)	−38(8)	−179(5)	0·15	0·45	24	Kinked	Kink at C terminus 82–94 = irregular
	100–118*	−63(6)	−42(9)	179(6)	0·10	0·15	184	Linear	
	124–149*	−63(6)	−43(7)	179(4)	0·15	0·57	88	Curved	Slight kink at C terminus
Actinidin	25–42*†	−63(8)	−44(6)	179(2)	0·11	0·32	78	Curved	Slight kink at C terminus
	69–80*	−61(8)	−42(11)	179(4)	0·10	0·39	31	Curved	
	120–130*	−63(3)	−38(8)	179(1)	0·09	0·30	34	Curved	
Crambin	6–17*	−62(4)	−42(4)	179(2)	0·09	0·30	60	Curved	
aPP	14–32†	−64(5)	−41(5)	179(3)	0·08	0·37	68	Curved	
Cytochrome *c*	2–13*	−63(5)	−41(7)	177(3)	0·10	0·27	48	Curved	
	87–101*	−65(4)	−39(5)	179(3)	0·08	0·20	94	Linear	
Flavodoxin	10–26	−65(6)	−39(8)	178(3)	0·11	0·16	137	Linear	Slight kinks at N and C termini
	93–106	−64(12)	−39(11)	180(3)	0·12	0·22	66	Curved	
	124–136*	−62(6)	−41(7)	177(3)	0·06	0·24	43	Curved	
Ovomucoid	34–43*†	−64(5)	−42(6)	179(4)	0·06	0·20	54	Curved	
Melittin	1–10*	−61(4)	−45(8)	180(2)	0·06	0·15	63	Curved	Slight kink at C terminus
	13–26†	−59(7)	−47(7)	180(2)	0·07	0·18	89	Curved	
Erythrocruorin	2–15	−67(6)	−37(6)	−179(7)	0·09	0·21	72	Curved	
	19–31	−64(9)	−40(6)	180(5)	0·10	0·22	63	Curved	
	52–72	−65(7)	−40(5)	179(7)	0·19	0·69	43	Kinked	Kink at Ile62
	76–87*	−64(10)	−40(6)	177(4)	0·07	0·26	80	Curved	Slight kink at C terminus
	93–111*	−64(7)	−40(8)	177(4)	0·10	0·24	112	Linear	Slight kink at C terminus
	117–135*	−65(8)	−41(11)	177(6)	0·16	0·47	50	Irregular	
Phospholipase	2–12*†	−64(4)	−41(5)	179(2)	0·05	0·26	53	Curved	
	39–55*	−64(3)	−40(5)	179(2)	0·13	0·16	112	Linear	
	89–107*	−62(7)	−43(5)	180(2)	0·13	0·30	71	Curved	
Dihydrofolate reductase	23–32*	−60(10)	−43(7)	180(4)	0·08	0·15	46	Curved	Slight kink at N terminus
	78–88*	−63(6)	−42(9)	179(4)	0·39	0·25	21	Irregular	
Mean values		−62(7)	−41(7)	179(3)	0·14(1)	0·29(1)	67(32)		

The 1st and last residues in each helix (listed in column 2) were initially defined using the DSSP program of Kabsch & Sander (1983), but were modified if HBEND indicated gross distortions at the termini. The helices that were shortened by 1 residue at the C terminus are indicated by*; those that were shortened by 1 residue at the N terminus are indicated by †.

⟨φ⟩, ⟨ψ⟩ and ⟨ω⟩ indicate the average values of the polypeptide main-chain torsion angles; figures in brackets show the corresponding standard deviations.

Each helix has been classified as linear, curved, kinked or irregular, according to the criteria summarized in Table 3. R_c is the mean deviation of all axis-points from the best circle fitted through the points, R_l is the corresponding measurement for a linear axis, and Radius refers to the radius of curvature of the helix axis. The positions of kinks in the helices are given in the column headed Comments.

As a result of this they have a slightly improved staggering of side-chains (see Fig. 3).

No π helices were observed in this 57 protein sample, but we note that one example has recently been discovered in catalase (Vainshtein *et al.*, 1986).

3. Analysis of the Regularity and Curvature of α-Helices

(a) *Definition of α-helices*

The dataset used for this analysis (Table 2) consisted of 48 α-helices, each of more than two turns in length, selected from 16 proteins with structures determined to a resolution of 1 to 2 Å.

The N- and C-terminal residues of each helix were initially identified using the secondary-structure definition program (DSSP) due to Kabsch & Sander (1983). However, if gross distortions of the helix termini were observed (usually because of residues with positive ψ angles, see below) these limits were modified. In the final analysis, 71% of the helices were shortened by one residue at the C terminus, and 25% were shortened by one residue at the N terminus, i.e. distortions at the C terminus of an α-helix are nearly three times more common than those at the N terminus (cf. Richardson, 1981).

(b) *Methods*

Initially, the analysis of α-helix regularity was attempted using plots showing the residue-by-residue variations in the helix dihedral angles ϕ and ψ, and n, the number of residues per turn. As outlined below, however, these data proved to be both qualitatively and quantitatively inadequate. An alternative strategy was therefore developed, using the purpose-built computer program HBEND. This program calculates the "axis" of each helix, and simultaneously provides data that can be used to classify it as regular or irregular, linear, curved or kinked.

The axis that is calculated represents the path followed by a model-built "probe" helix as a moving fit is obtained for this helix along the length of the "real" helix. Different probe helices are used to analyse different real helices, with the values of ϕ, ψ and ω for each probe helix taken as the mean values of the real helix analysed. (The alternative approach, which involves fitting a "standard" probe helix to each real helix, generates an axis that precesses about the "true" real helix axis, if the radius and pitch of the 2 helices are significantly different.)

The algorithm employed by the program (briefly described by Blundell *et al.* (1983)) is summarized below:

For each residue (i) the probe helix and real helix are superposed to give a least-squares fit for atoms C'_{i-1}, N_i, C^{α}_i, C'_i, N_{i+1}. After each fit, the co-ordinates are calculated for the projection of the (real helix) atom C^{α}_i, on to the probe helix axis. The set of points thus generated, describes the course of the real helix axis (see for example Fig. 4).

Calculation of a radius of curvature for the helix involved:

(1) calculating the least-squares plane through the points;

(2) projecting the points on to this plane;

(3) fitting the least-squares circle through the projected points; and

(4) calculating the radius of this circle.

The r.m.s.† deviation of the points from the circle (R_c, in Å) gives a measure of how well the axis is described by a curve. The corresponding quantity calculated for the least-squares straight line fitted through the points (in 3-dimensional space; R_l, in Å), provides the same data for a linear axis.

Additional data provided by HBEND include: the mean and standard deviations of all real helix

† Abbreviations used: r.m.s., root-mean-square; r.m.s.e., root-mean-square error.

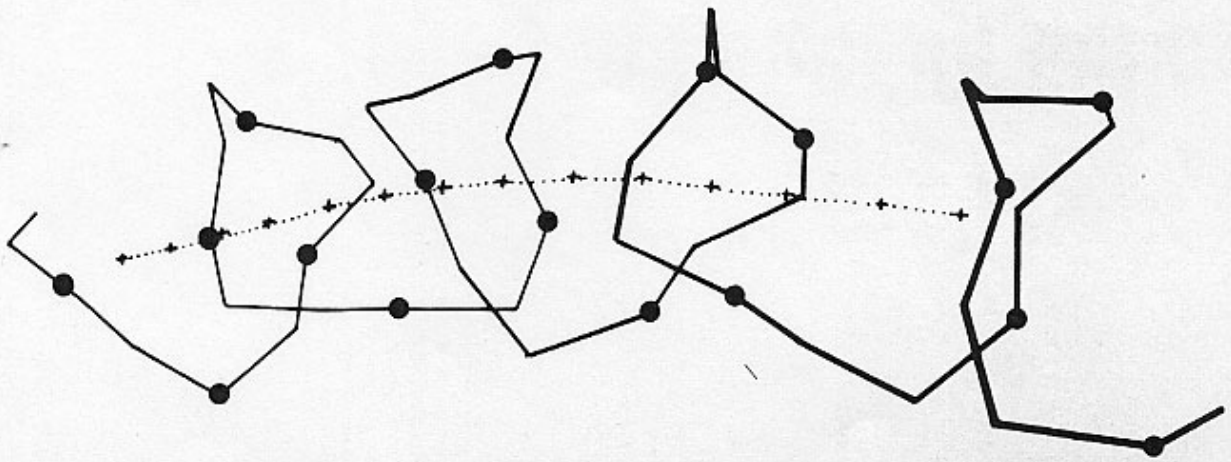

Figure 4. A curved helix involving residues 14 to 32 in avian pancreatic polypeptide, showing the "axis" generated by the HBEND program. The axis is represented by a series of broken lines joining consecutive "axis-points" (shown as +). The helix is viewed approximately perpendicular to the axis, and only the main-chain atoms are indicated. C^{α} atoms are marked by filled circles.

Table 3

A summary of the criteria used to distinguish between linear, curved, kinked and irregular α-helices

	Radius of curvature	R_l	R_c	r.m.s.e.
Linear	>90 Å	≤0·25 Å	≤0·15 Å	—
Curved	≤90 Å	—	≤0·15 Å	—
Irregular	Low (Usually ≤45 Å)	$\leq R_c$	>0·15 Å	—
Kinked	Variable (Usually ≤45 Å)	Usually ≫0·25 Å	—	Kink at site (i) if r.m.s.e.$_i$ ⟩ ⟨r.m.s.e.⟩ + $2\sigma_r$

For details of the symbols used, refer to the text and the legend for Table 2.

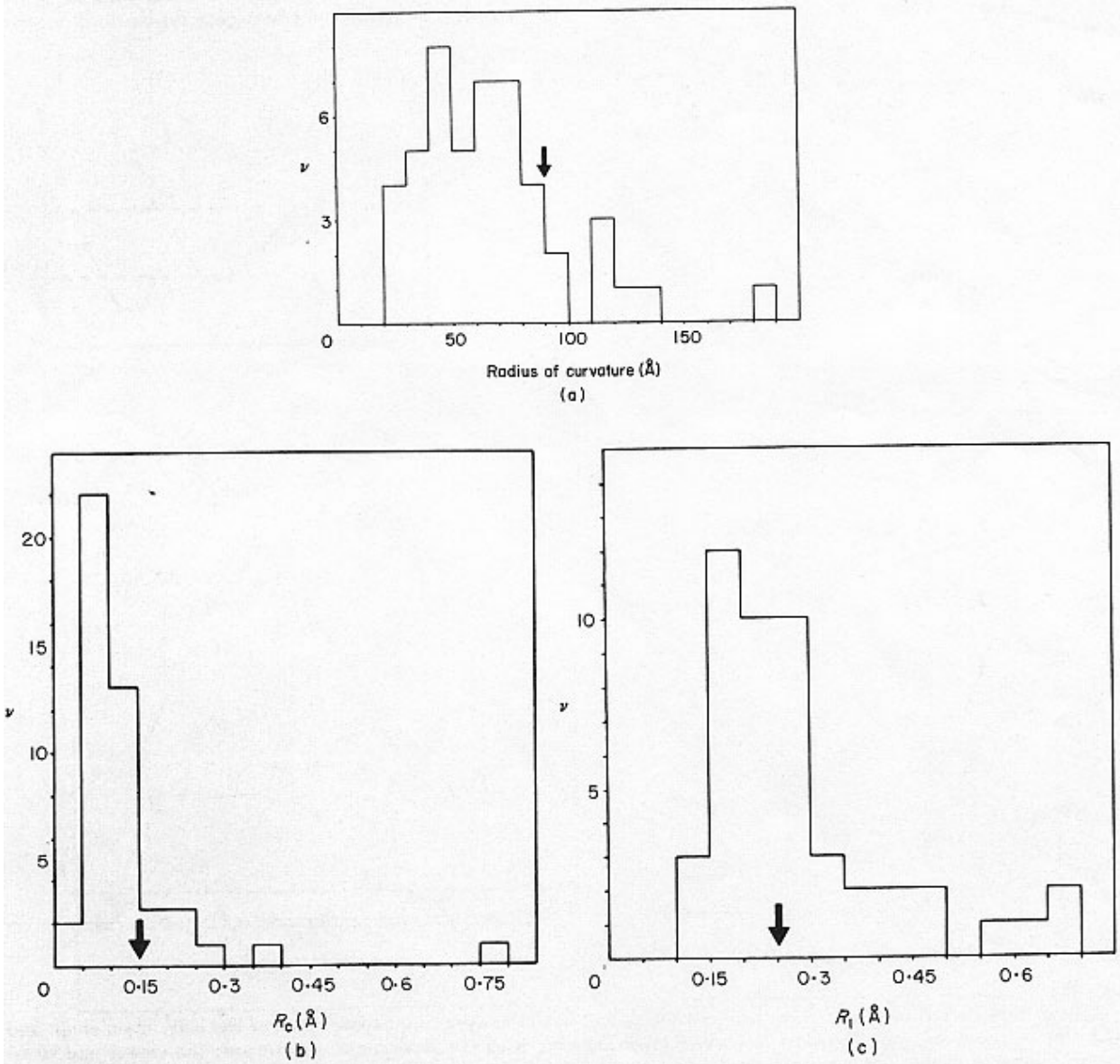

Figure 5. Helix curvature and regularity. (a) The distribution of the radius of curvature for the observed α-helices. (b) The distribution of R_c for the observed α-helices, where R_c is the mean distance of all axis-points from the best circle fitted through them. (c) The distribution of R_l for the observed α-helices, where R_l is the mean distance of all axis-points from the best straight line fitted through them. The radius of curvature, R_c and R_l are all measured in Å.

main-chain torsion angles ($\langle\phi\rangle$, σ_ϕ etc.), the root mean square error (r.m.s.e., in Å) for each fit of the probe helix to the real helix, and the mean and standard deviation of all r.m.s.e. values (respectively $\langle$r.m.s.e.$\rangle$ and σ_r). The program also issues warnings to indicate residues that have abnormal main-chain torsion angles, and those fits of the probe helix and real helix for which the r.m.s.e. is greater than $\langle$r.m.s.e.$\rangle + 2\sigma_r$.

(c) *Results*

The results of the HBEND-analysis are presented in Table 2. Each helix is classified as linear, curved, kinked or irregular, according to the criteria that are described below, and summarized in Table 3. Although these criteria are, to some extent, arbitrary they facilitate the description of the different types of helix distortion.

The structures and HBEND data for the different classes of helix are illustrated in Figures 6 to 9. (For comparison, we also show the corresponding variations in ϕ, ψ and n).

(d) *Regular helices*

Thirty-five of the helices analysed (73% of the total sample) are classified as linear or curved. These two types of helix are jointly referred to as regular helices, and are distinguished from one another on the basis of their radius of curvature, and the values of R_l and R_c.

R, the radius curvature of a helix, shows a fairly continuous distribution over the range 21 to 184 Å, with a broad peak centred near the mean radius of 67 Å (Fig. 5(a)). From a consideration of the overall distribution, a value of $R < 90$ Å was chosen as the cut-off to distinguish between linear and curved

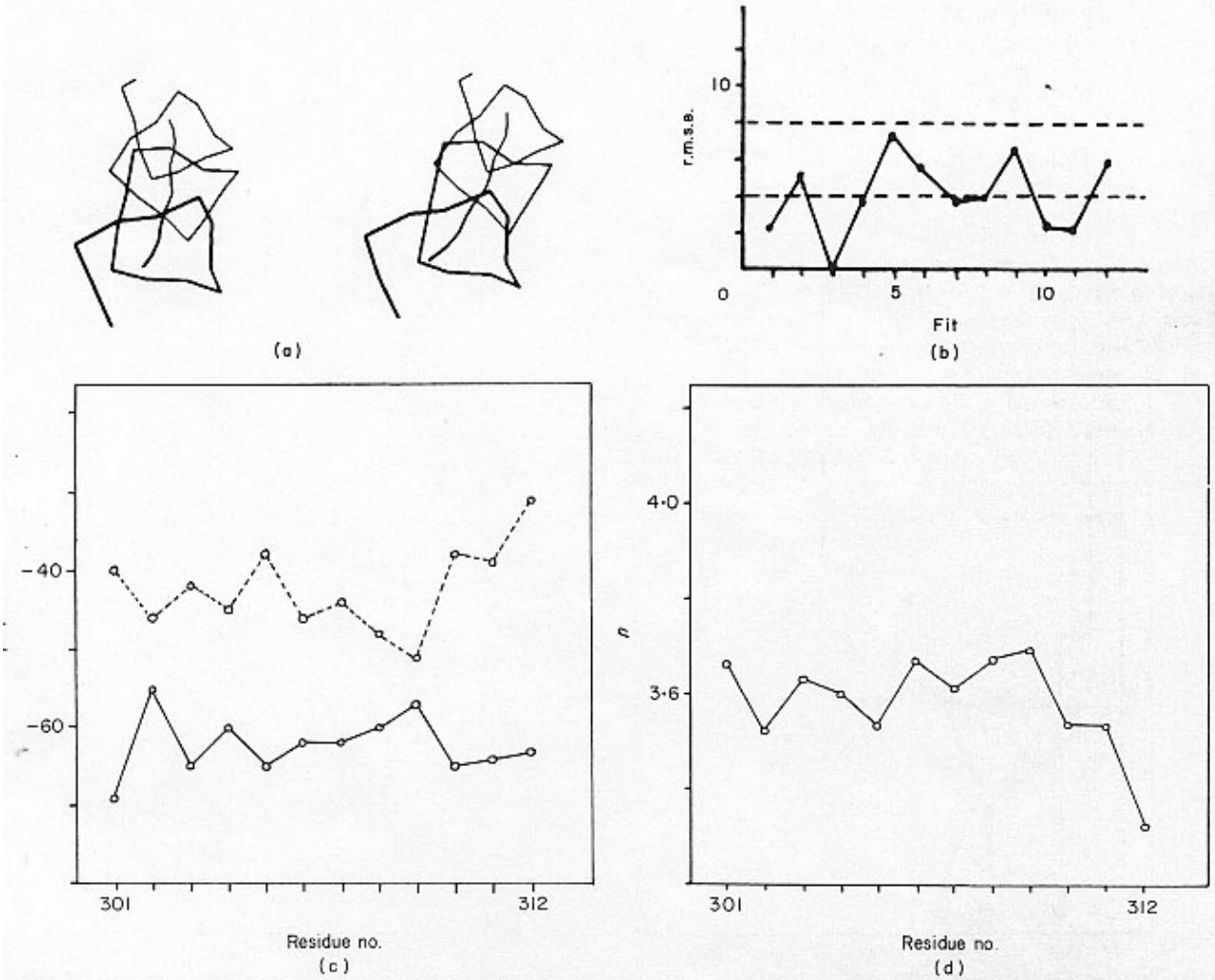

Figure 6. A curved helix involving residues 300 to 313 in thermolysin. (a) Stereo-views of the helix main-chain and the path of the helix axis. The helix is viewed down the axis, with the N-terminal end towards the viewer and tilted slightly downwards. (b) The corresponding r.m.s.e. plot (with r.m.s.e. values in units of Å $\times 10^{-3}$). Continuous lines show the variation in r.m.s.e. values, broken lines show $\langle$r.m.s.e.$\rangle$ and $\langle$r.m.s.e.$\rangle + 2\sigma_r$. (r.m.s.e. values represent the root-mean-square error for each fit of the probe helix to the helix analysed, see Analysis of the Regularity and Curvature of α-Helices, section (b).) (c) and (d) The residue-by-residue variations in the dihedral angles ϕ and ψ (in degrees), and n, the number of residues per turn.

helices. The majority of helices (83%) have R <90 Å. Depending upon the values of R_c and R_l (see below), helices with R<90 Å were taken as curved helices, and those with R>90 Å were considered to be linear.

From the distributions of R_c and R_l (Fig. 5(b) and (c)), it is clear that most of the helices have an axis that is better described as a curve, rather than a straight line; a value of R_c=0·15 Å was chosen as the upper limit for curved helices.

The majority (28) of the regular helices are classified as curved. Since the r.m.s.e. plots for these helices are generally featureless (see for example Fig. 6), it is clear that they "bend" gradually over their entire length.

In each of the curved helices, the centre of curvature lies on the hydrophobic side of the helix, roughtly on a line defined by the "hydrophobic moment" (Eisenberg *et al.*, 1982). Examples of curved helices are shown in Figures 4 (avian pancreatic polypeptide) and 6 (thermolysin).

Although seven of the curved helices also have a slight kink at the N or C terminus (usually the latter), this is not the main cause of their curvature: if the HBEND analysis is repeated with each helix shortened by one residue at the appropriate end, only one of the seven would no longer be considered curved. This helix, involving residues 234 to 246 in thermolysin, would, in its shortened form (residues 234 to 245) be classified as irregular.

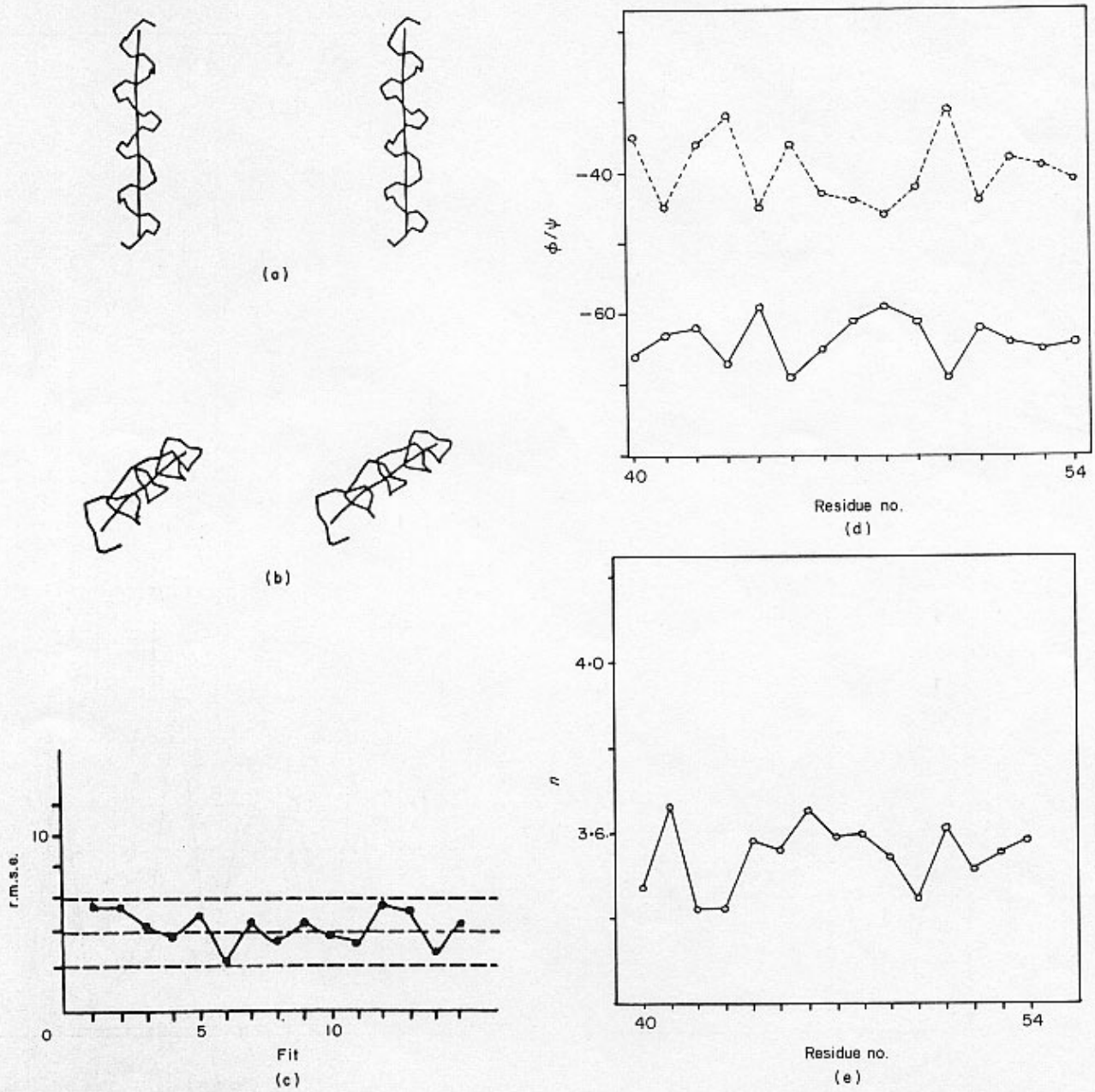

Figure 7. A linear helix involving residues 39 to 55 in phospholipase. (a) and (b) show stereo-views of the helix main-chain and the path of the helix axis; (c), (d) and (e) correspond to (b), (c) and (d) in Fig. 6. For further details of the presentation, see the legend to Fig. 6.

Only seven of the regular helices (15% of the total sample) are classified as "linear". Figure 7 shows the "linear" helix found in phospholipase, which has a radius of curvature of 112 Å.

(e) *Non-regular helices*

Thirteen of the helices analysed are considered to be non-regular. They are sub-divided into those that have kinks in the middle of their structure, those that have kinks at the N and/or C terminus, and those that are distorted at several points along their length.

The five helices that fall into the latter category are listed in Table 2 as irregular. These helices all have $R_c > 0{\cdot}15$ Å, and four of them also have $R_c \geq R_l$. They are thus neither linear, nor smoothly curved. Figure 8 shows an irregular helix found in myoglobin.

The remaining eight non-regular helices are classified as kinked. Four of these helices have kinks in the middle of their structure, and four have kinks at their termini. Both types of kink are identified from the plot of the helix r.m.s.e. values. Since these values reflect the variations in radius and pitch of the helix, any abrupt changes in the main-chain conformation give rise to fits for which r.m.s.e. $> \langle$r.m.s.e.$\rangle + 2\sigma_r$. This is illustrated for the case of the E-helix in erythrocruorin (Fig. 9), where a kink is observed in the region of Ile62.

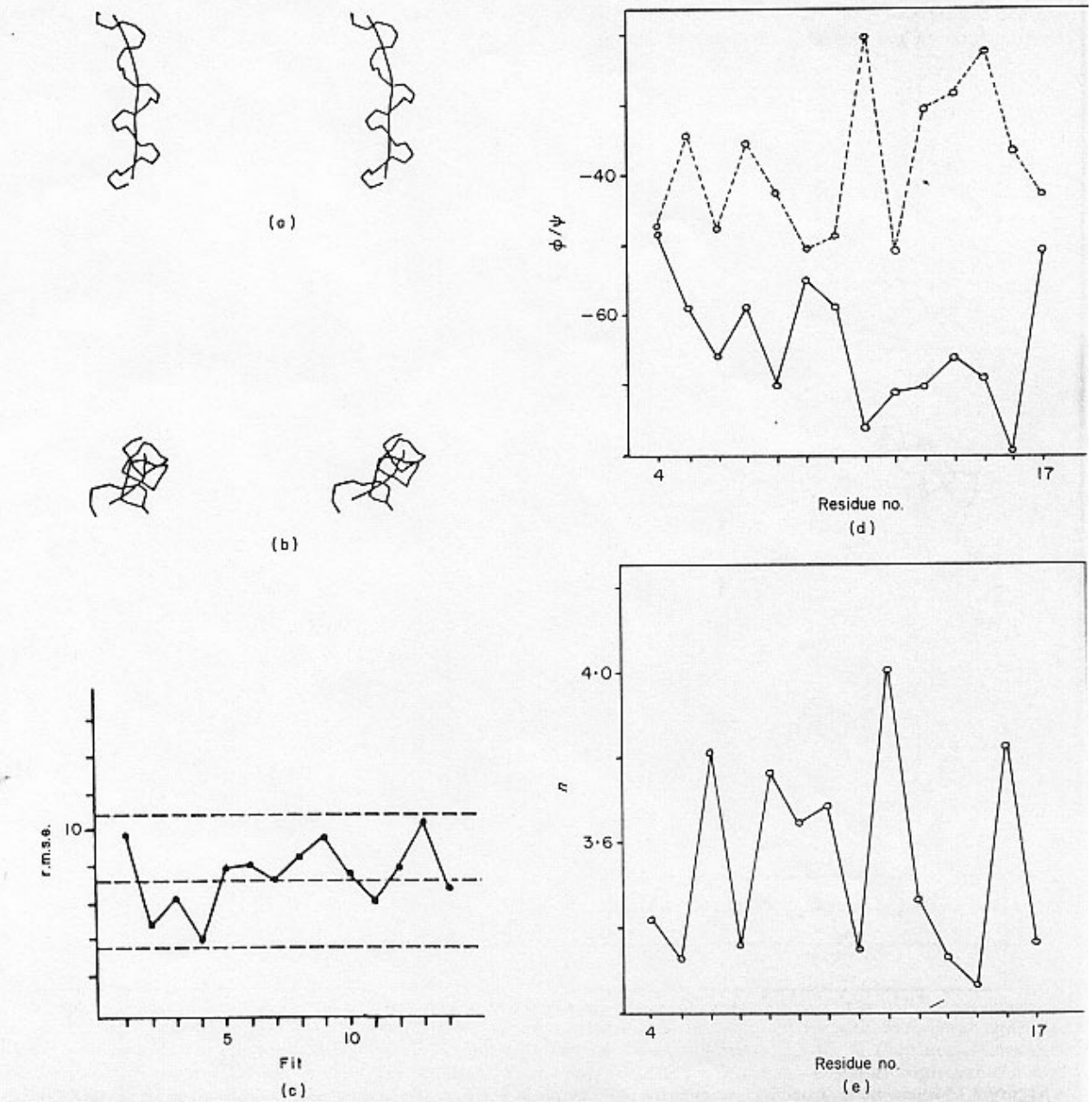

Figure 8. An irregular helix involving residues 3 to 18 in myoglobin. For details of the presentation, see the legend to Figs 6 and 7.

Fitting the best straight line axes to the two "halves" of the helix (residues 52 to 62 and 62 to 72) shows that they are related by an angle of 20°.

Other helices that have kinks in the centre of their structure include: residues 136 to 151 in thermolysin, where a kink occurs near Val139, residues 159 to 180 in thermolysin, where a kink occurs near Glu166, and residues 72 to 89 in carboxpeptidase, where a kink occurs near Thr75. The angles of kink for these helices are, respectively, 31°, 30° and 20°.

As for the helices that have kinks at their termini, three are distorted only at the C terminus, and one is distorted at both termini. The r.m.s.e. plot for this type of helix is illustrated for the case of the H helix in myoglobin (Fig. 10). It can be seen that there is little variation in r.m.s.e. for fits 1 to 23 of the probe helix, but a sharp increase in r.m.s.e. for fit 24.

(f) *Intrahelix variations in ϕ, ψ and* n

For regular helices (i.e. curved and linear) the ϕ and ψ angles show an inverse correlation, so that increases in ϕ are matched by decreases in ψ, and *vice versa* (see Figs 6(c) and 7(d)). This approximate symmetry maintains the regularity of a helix, and is not seen in the ϕ/ψ plots for irregular helices (Figs 8(d) to 9(d)). This is also reflected in the plots showing variations in n (Figs 6(d) and 7(e) to 9(e)).

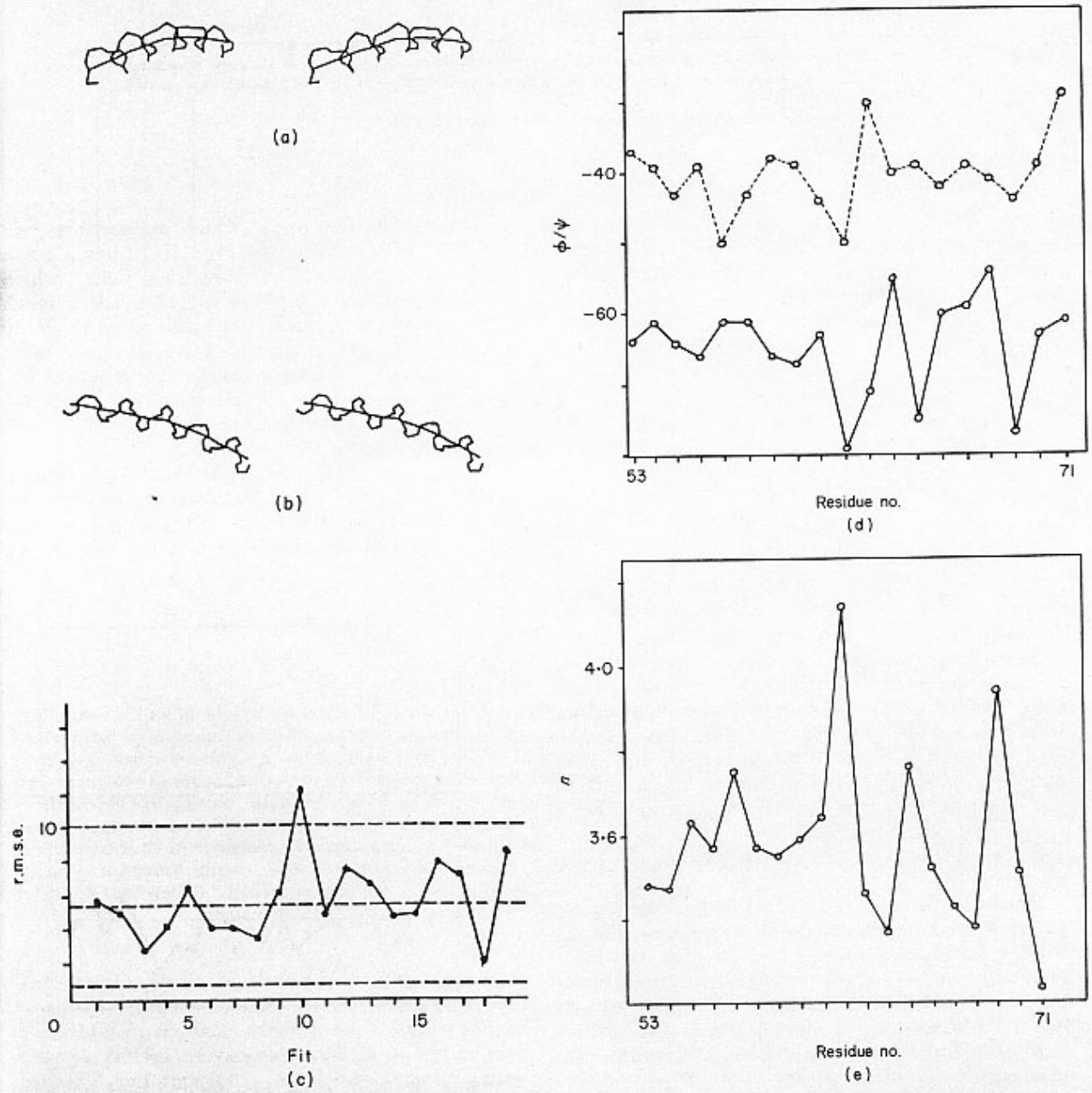

Figure 9. A kinked helix involving residues 52 to 72 in erythrocruorin. For details of the presentation refer to the legend to Figs 6 and 7.

Table 4
A comparison of the structures of proline-kinked α-helices

Protein	Pro	Helix	r.m.s. (Å)	θ	Local sequence
Sea lamprey haemoglobin	25	14–30	—	26°	RSAWA P VYSDY
Adenylate kinase	159	142–168	1·05	19°	YKATE P VIAFY
Glutathione reductase	65	55–80	0·86	28°	NVGCV P KKVMW
Citrate synthase	15	5–27	0·36	28°	LADLI P KEQAR
	183	166–195	0·40	26°	LIAKL P CVAAK
Myoglobin	88	82–96	0·48	21°	EAELK P LAQSH
Liver alcohol dehydrogenase	329	324–336	0·36	24°	SKDSV P KLVAD
Cytochrome peroxidase	94	86–98	0·37	29°	FKFLE P IHKEF
Glyceraldehyde-3-phosphate dehydrogenase	G156	G148–G166	0·23	35°	TNCLA P VAKVL
Melittin	A14	A1–A26	0·37	49°	LTTGL P ALISW

r.m.s. is the root-mean-square error obtained when each helix is superposed on to the helix involving residues 14 to 30 in sea lamprey haemoglobin, to give a least-squares fit for the 7 α-carbon atoms of residues $i-3 \rightarrow \text{Pro}(i) \rightarrow i+3$.
θ is the the angle of kink in each helix (as defined in Fig. 11(a)).

These observations can thus be used to distinguish between regular and irregular helices (cf. the use of $(\phi+\psi)$ plots by Steigemann & Weber (1979)), but do not distinguish between linear and curved helices, or between kinked and irregular ones. Indeed for the kinked helices these parameters do not even permit an unambiguous assignment of the kink site: for the example illustrated (Fig. 8), there are major changes in both the dihedral angles and n for the kink residue, Ile62, but comparable changes are also seen for residues 64 and 69.

(g) *The effects of proline on an α-helix*

The early statistical analyses of protein secondary structures, carried out first of all by Chou & Fasman (1974), and later by Levitt (1978), demonstrated that, while proline residues are often found at the ends of an α-helix, they seldom occur in the middle of their structure. These observations were rationalized by supposing that proline residues lead to an unacceptable disruption of the helix hydrogen bonding, and that by dictating a kink in the structure, they also disturb the packing of helix side-chains. As a consequence of this there is a widespread belief that proline residues act as α-helix "breakers".

However, in a survey of the 291 helices in 57 proteins, we have discovered ten examples that do contain an internal proline (see Table 4). We have analysed the sequences and structures of these helices, firstly to examine the extent of the distortion caused by proline, and secondly to find out how and why these helices tolerate the distortion.

As shown in Table 4, nine of the ten helices kink by roughly the same amount. For these nine structures, θ, the angle which relates the two "halves" of the helix (see Fig. 11(a)), has a mean value of $26(\pm 5)°$. The remaining helix, which involves residues 1 to 26 in melittin, has a much more pronounced kink (with $\theta=49°$), probably because this allows the monomers to pack more efficiently in the tetramer (Terwilliger & Eisenberg, 1982). (Note that for the purpose of the analysis described in the preceding section, this helix was considered as 2 separate helices.)

The conformational changes that take place to accommodate the proline vary from one structure to another. In most cases the changes are relatively minor, so that all residues retain the ϕ/ψ angles

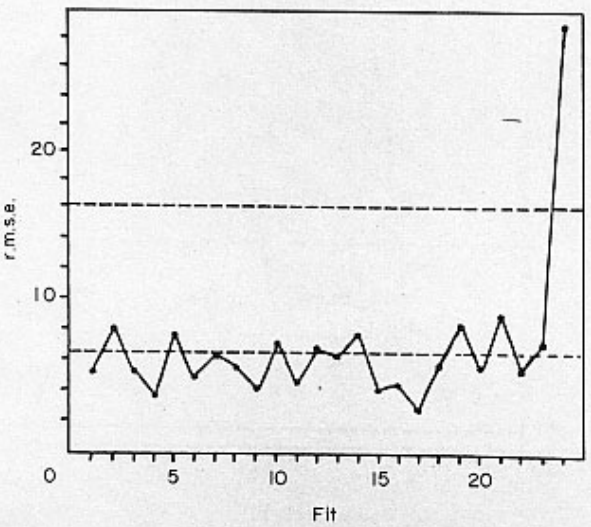

Figure 10. The r.m.s.e. plot for the helix involving residues 124 to 149 in myoglobin. The helix is kinked at the C terminus.

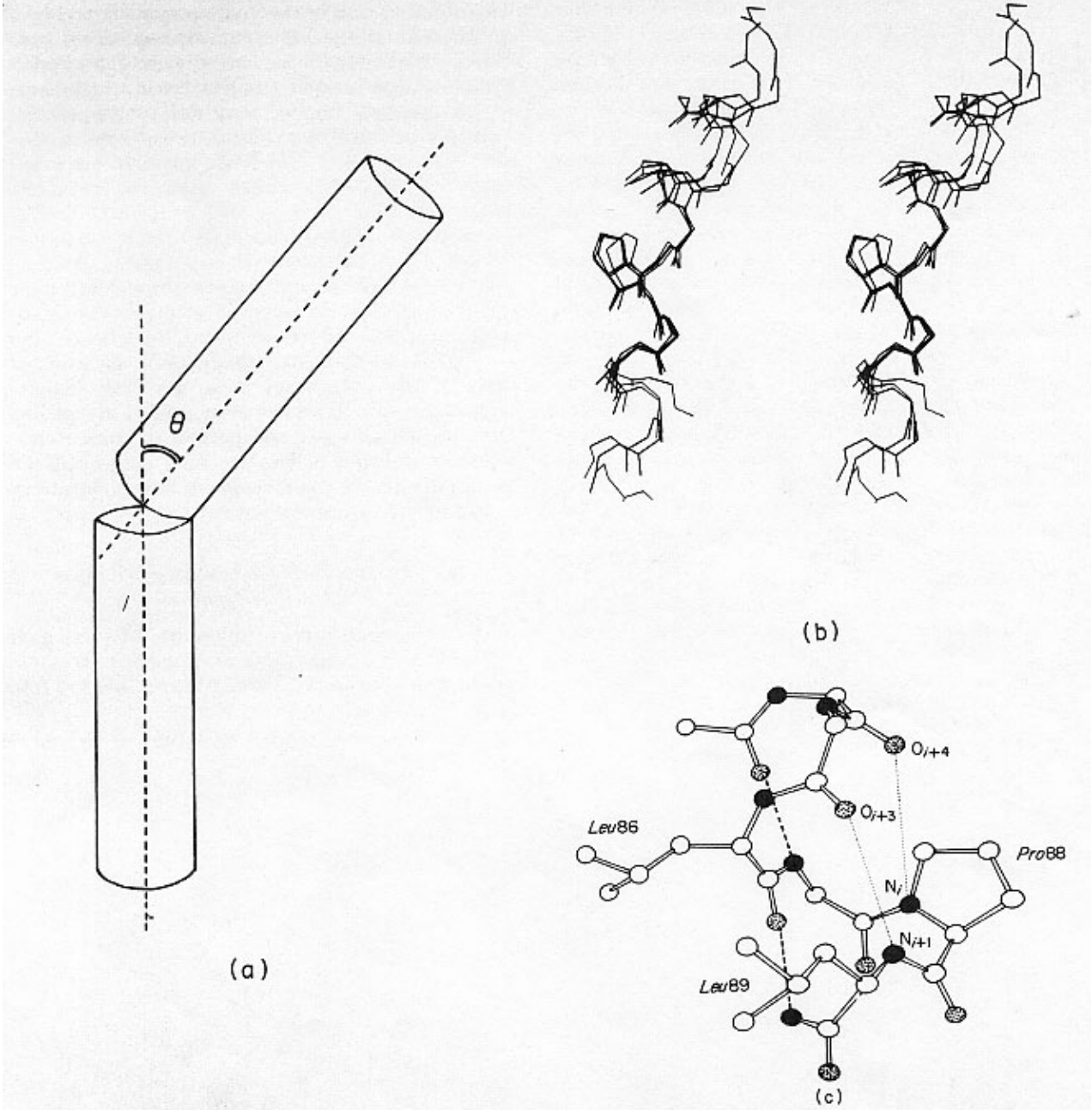

Figure 11. Proline-kinked α-helices. (a) A schematic diagram showing how the kink angle (θ) is defined. The α-helix is represented by a broken cylinder, and the axes of the N- and C-terminal "halves" are shown as broken lines. (b) Four proline-kinked helices superposed to give a least-squares fit for the (7) C^{α} atoms of the residues $i-3\rightarrow$(Pro)$i\rightarrow i+3$. The 4 helices involve residues 14 to 30 in sea lamprey haemoglobin, residues 166 to 195 in citrate synthase, residues 324 to 336 in alcohol dehydrogenase, and residues 86 to 98 in cytochrome peroxidase. (c) An Ortep plot (Johnson, 1976) showing how a proline residue disrupts the pattern of α-helical hydrogen bonding. The section of helix illustrated involves residues 83 to 89 in myoglobin. Oxygen atoms are shown as stippled ellipses, nitrogen atoms as filled ellipses and carbon atoms as unshaded ellipses. "Lost" hydrogen bonds are indicated by dotted lines and existing ones by broken lines. The length of the "broken" hydrogen bond, $O_{i-3}-N_{i+1}$ is 4·1 Å.

appropriate for a right-handed helix. The two exceptions are the helices found in melittin and glutathione reductase, where there are more radical changes caused by residues with β- or left-handed helical conformation. A least-squares superpositioning of the ten helices gives a mean r.m.s. error of 0·45 Å, for a fit of the (7) C^{α} atoms of the residues $i-3\rightarrow$(Pro)$i\rightarrow i+3$. Figure 11(b) shows the superpositioning of four of the helices.

All ten of the helices show surprisingly little disruption of the hydrogen bonding. Each one of course lacks the hydrogen bond: $N_i\cdots(O_{i-4})$, because the proline (i) has no amide proton. However, the only hydrogen bond that is actually

"broken" as a result of the kink, is the one involving atoms N_{i+1} and O_{i-3} (see Fig. 11(c)). The three groups that are not hydrogen bonded within the helix (O_{i-3}, O_{i-4}, N_{i+1}) are either in a position to be able to hydrogen bond to solvent, or are hydrogen bonded to other groups within the protein. Given that the loss of two hydrogen bonds will upset the co-operativity required for helix formation, it is not surprising to find that the proline-kinked helices are all long helices (Table 4); the helix lengths range from four turns to eight turns and the mean length is just over five turns. Although the sequences of the helices lack the normal patterns of hydrophobic (H) and hydrophilic (h) residues (e.g. HHhhHH etc.), there appears to be no pattern that could be said to be characteristic of a proline-kinked helix. We conclude that the kinks in these helices are caused solely by the presence of the prolines. These residues have been conserved during the course of protein evolution. For example, the proline that occurs at position 183 in the porcine sequence of citrate synthase is also found in the enzymes from yeast (Pro223) and *Escherichia coli* (Pro1170); the homology between the porcine and yeast enzymes is 62%, and between the *E. coli* and yeast enzymes only 27%. The same degree of conservation is also seen in glyceraldehyde-3-phosphate dehydrogenase, where proline 156 in the yeast enzyme is conserved in all eight of the other known sequences, which derive from organisms as diverse as *Bacillus stearothermophilus* and humans. Such a high degree of conservation implies that the proline residues have a definite structural/functional role in these proteins, and that the kink in each helix is a necessary distortion, rather than an undesirable one.

(Note that although there are multiple sequences available for adenylate kinase, alcohol dehydrogenase and myoglobin, these sequences are either difficult to align because of problems associated with residue insertions and deletions (e.g. *Drosophila* alcohol dehydrogenase), or else are derived from organisms that are less distantly related (e.g. the 70 known sequences of myoglobin). However, in all cases the proline residues responsible for kinked α-helices are fully conserved. The examples that we describe above best illustrate the extent of their conservation.)

(h) *The interaction between an α-helix and solvent/side-chains*

To investigate the influence of carbonyl-solvent/side-chain interactions on α-helix structure, a survey was made of several α-helices selected from

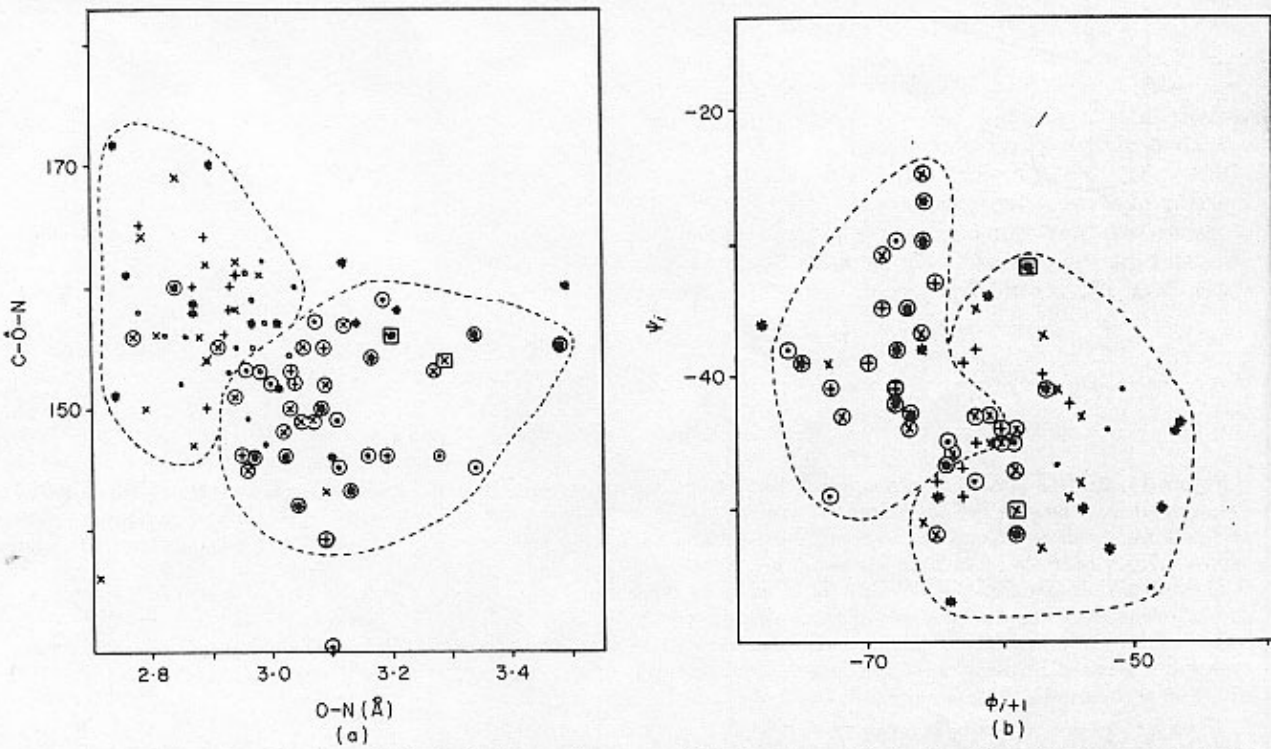

Figure 12. The effect of solvent/side-chain interactions on α-helix structure. (a) A plot of the C–Ô–N angles *versus* O–N distances for 12 α-helices from the proteins: actinidin, shown by × (the helices involving residues 25 to 43, 71 to 81 and 121 to 129); ribonuclease, shown by ● (the helices involving residues 3 to 13 and 24 to 33); avian pancreatic polypeptide, shown by + (the helix involving residues 15 to 32); dihydrofolate reductase, shown by * (the helices involving residues 23 to 35, 42 to 49, 78 to 89 and 99 to 106); porcine insulin (Sakabe *et al.*, 1981) shown by ○ (the helices involving residues B9 to B19 and D9 to D19). Where CO groups are also hydrogen bonded to a water molecule or side-chain group, the symbols are circled. (Symbols that are boxed indicate the data for bifurcated hydrogen bonds.) (b) A graph of the torsion angles ψ_i and ϕ_{i+1} (in degrees), for the α-helices listed above. (Symbols used are the same as in (a).)

highly refined protein structures. The preliminary results from the analysis (considering only 6 α-helices) were reported by Blundell *et al.* (1983), and have subsequently been supported by the work of Baker & Hubbard (1984). For the sake of completeness, we include here the results obtained using a larger sample of helices, but in view of the previous literature, present only limited details. (The 12 helices considered in the analysis are listed in the legend to Fig. 12.)

The analysis demonstrates that the helix hydrogen bonds that involve CO groups that are also hydrogen bonded to solvent or side-chain groups, are longer and less linear than those formed by CO groups on the "buried" side of an α-helix (see Fig. 12(a)). This is because the CO groups that also hydrogen bond to solvent or side-chains are tilted further outwards from the helix axis (an observation first recorded by Watson (1969); Fig. 13). The mean O–N distances for "hydrophilic" and "buried" hydrogen bonds are, respectively, 3·09 (±0·13) Å and 2·91 (±0·06) Å; the corresponding mean C–Ô–N angles for the two classes are 148 (±6)° and 157 (±5)°.

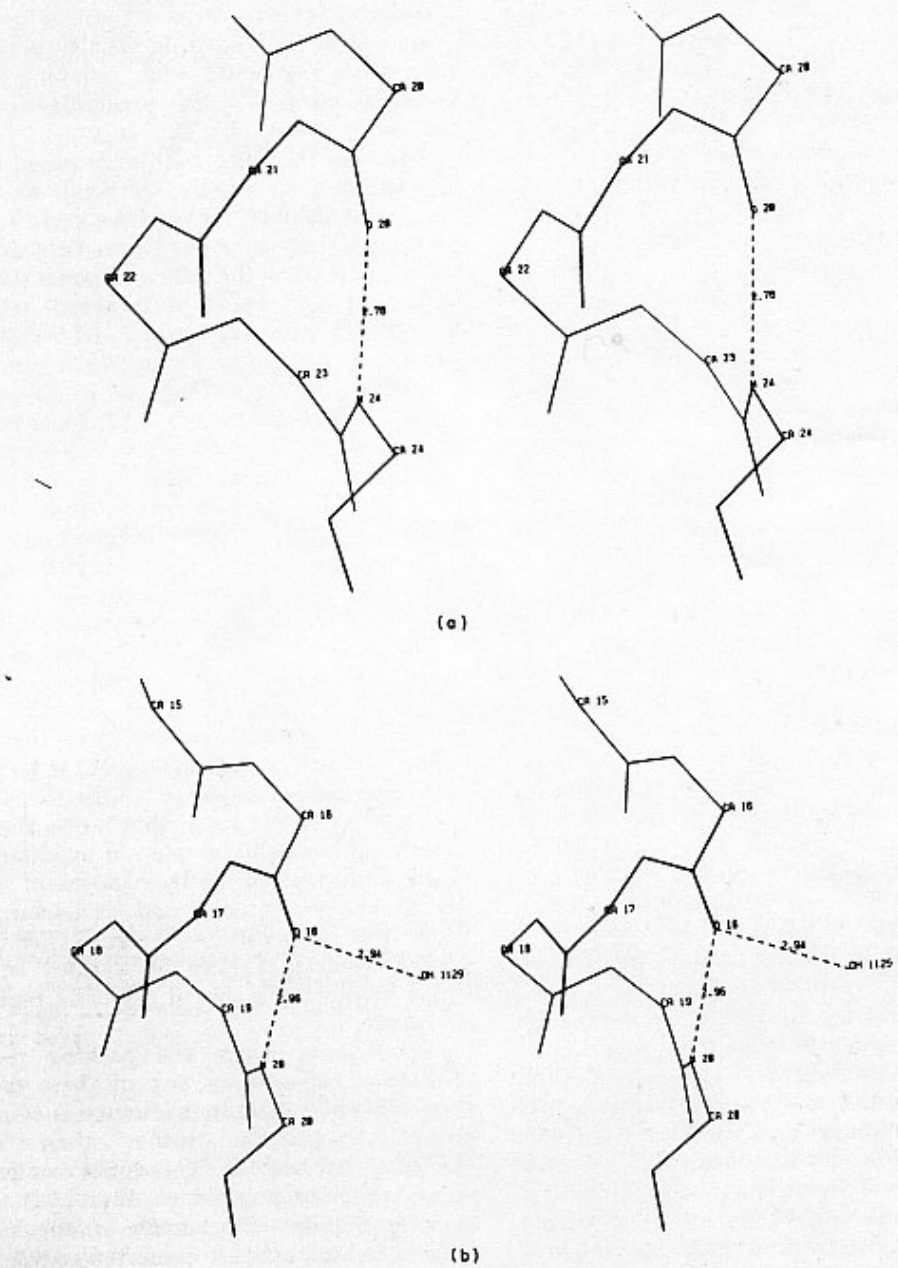

Figure 13. Stereo-views of α-helix hydrogen bonds in avian pancreatic polypeptide. The diagrams illustrate the decreased C-Ô-N angle and increased O–N distance, for CO groups hydrogen bonded to a water molecule. Hydrogen bonds are shown as broken lines. O–N distances are indicated in the diagrams. CA is C^{α}. (a) The CO group of residue 20, where the oxygen atom is hydrogen bonded only to the NH group of residue 24. The C–Ô–N angle is 165°. (b) The CO group of residue 16, where the oxygen atom is hydrogen bonded to the NH group of residue 20, and also to a water molecule (shown as OH). The C–Ô–N angle is 146°.

This variation in tilt of the helix carbonyl groups is reflected in the helix main-chain torsion angles: ψ_i and ϕ_{i+1} (Fig. 12(b)). For all residues i that have hydrophilic hydrogen bonds, the mean ψ_i, ϕ_{i+1} angles -41 $(\pm 6)^\circ$, -66 $(\pm 5)^\circ$, and the corresponding mean values for residues with "buried" hydrogen bonds are -44 $(\pm 6)^\circ$, -59 $(\pm 6)^\circ$.

(i) *Summary and discussion*

The results presented in this paper show that up to 85% of α-helices are in some way distorted. Out of the 48 helices studied, five are malformed throughout their length, eight are kinked and 28 are curved. Only seven helices (15% of the total sample) are "linear". The different types of helix distortion can be attributed to different factors:

(1) Solvent/side-chain interactions.

(2) The local sequence (in particular proline residues).

(3) Side-chain packing.

In the case of regularly curved (amphipathic) α-helices the distortions are attributed to hydrogen bond effects. We have shown that by comparison with the residues on the buried side of an α-helix, those on the (solvent) exposed side have their carbonyl groups tilted further outwards from the helix axis. This allows the exposed groups to satisfy their full hydrogen bond potential, because they can then accept one hydrogen bond from a helix amide, and another from the solvent or side-chains. For the carbonyl groups of residues on the buried side, however, this is not possible; these groups can accept only the one (α-helical) hydrogen bond. The buried hydrogen bonds are more linear than those on the exposed side, presumably to allow both of the lone pairs of electrons on the carbonyl oxygen atom to interact with the one amide proton.

The net effect of these differences in carbonyl orientation is to cause the residues on the two sides of a helix to have different main-chain conformations; those on the buried side have $(\phi_{i+1}, \psi_i) = (-59^\circ, -44^\circ)$, and those on the exposed side have $(\phi_{i+1}, \psi_i) = (-66^\circ, -41^\circ)$. This translates to a "bending" or curvature of the helix, because it means that the coils of the helix are compressed on the buried side and expanded on the exposed side.

These observations are supported by the results of Baker & Hubbard (1984) in their analysis of hydrogen bonds in proteins, and by those of Yang *et al.* (1969), in their study of water–carbonyl interactions in peptide crystals. In addition, it has been demonstrated by Chakrabati *et al.* (1986) that there are systematic differences in the peptide bond angles (O–$\hat{C}'$–N and N–$\hat{C}'$–C^α) for the buried and exposed residues in an α-helix, which can also lead to significant curvature; these authors show that a difference of 4° in the O–$\hat{C}'$–N bond angle between the two sides of an amphiphilic α-helix is sufficient to produce a radius of curvature of 70 Å.

For the curved helices analysed here, the radius of curvature is generally about 60 Å, which corresponds to ~1° of arc per residue. By comparison, we find that the radius of curvature for a heptapeptide α-helix in the coiled-coil conformation (calculated using co-ordinates provided by Parry & Suzuki (1969)), is ~150 Å, which corresponds to ~0·5° of arc per residue. (Note, therefore, that according to the criteria employed in this analysis, the coiled-coil α-helix, is linear!) If the ΔG for the deformation of an α-helix to the coiled-coil state is of the order of 0·1 kcal/mole per residue (as suggested by the energy calculations of Crick (1952) and Parry & Suzuki (1969)), then the distortion of a five turn helix from the linear to curved state, will involve an energy penalty of <2 kcal (1 cal = 4·184 J). That this penalty is so small (less than that associated with breaking a hydrogen bond) accounts for the predominance of curved helices.

Although the perturbations required to cause the curvature of an α-helix are small at the level of individual residues, they can be greatly amplified in a long helix. As shown in Figure 14, this means that there is a marked difference between the end-points of equal-length linear and curved helices: if two four-turn helices are superposed to give a least-squares fit for the first two turns, the axis-to-axis separation at their C termini is about 4 to 5 Å. Thus, the relative positions of the side-chains in a curved α-helix, are quite different from those in a linear helix. This may be of benefit during protein evolution, since mutations that lead to changes in side-chain volume can be accommodated by minor changes in the helix main-chain conformation, e.g. a linear helix may become curved (or a curved helix, more curved) in order to preserve an optimal packing of side-chains. It can be seen in Table 2, for example, that the B helix in erythrocruorin (residues 19 to 31) is quite different from the B helix in myoglobin (residues 20 to 36): in erythrocruorin the helix is curved, and in myoglobin it is irregular. It has been suggested by Chothia (1984) that the pressures on an α-helix that arise from packing effects will normally be relieved by changes in side-chain conformation, or by changes in the relative orientation of the helix and its packing partners. From the observations recorded here, it would appear that there may be a third contribution, which involves changes in the helix regularity and curvature.

In instances where the packing requirements cannot be satisfied by any of these mechanisms, they are likely to produce some local and fairly abrupt irregularity, rather than a smooth "bending" of the helix. This is the case for the helix in erythrocruorin (residues 52 to 72), where the packing requirements dictate a 20° kink in the helix. The kink arises because the orientation of the helix preferred for the packing of side-chains at its interface with the A and H helices conflicts with that required for its packing against the B helix (Lesk & Chothia, 1980). The fact that the kink occurs near Ile62 is probably because this residue makes contact with the haem group, and in this

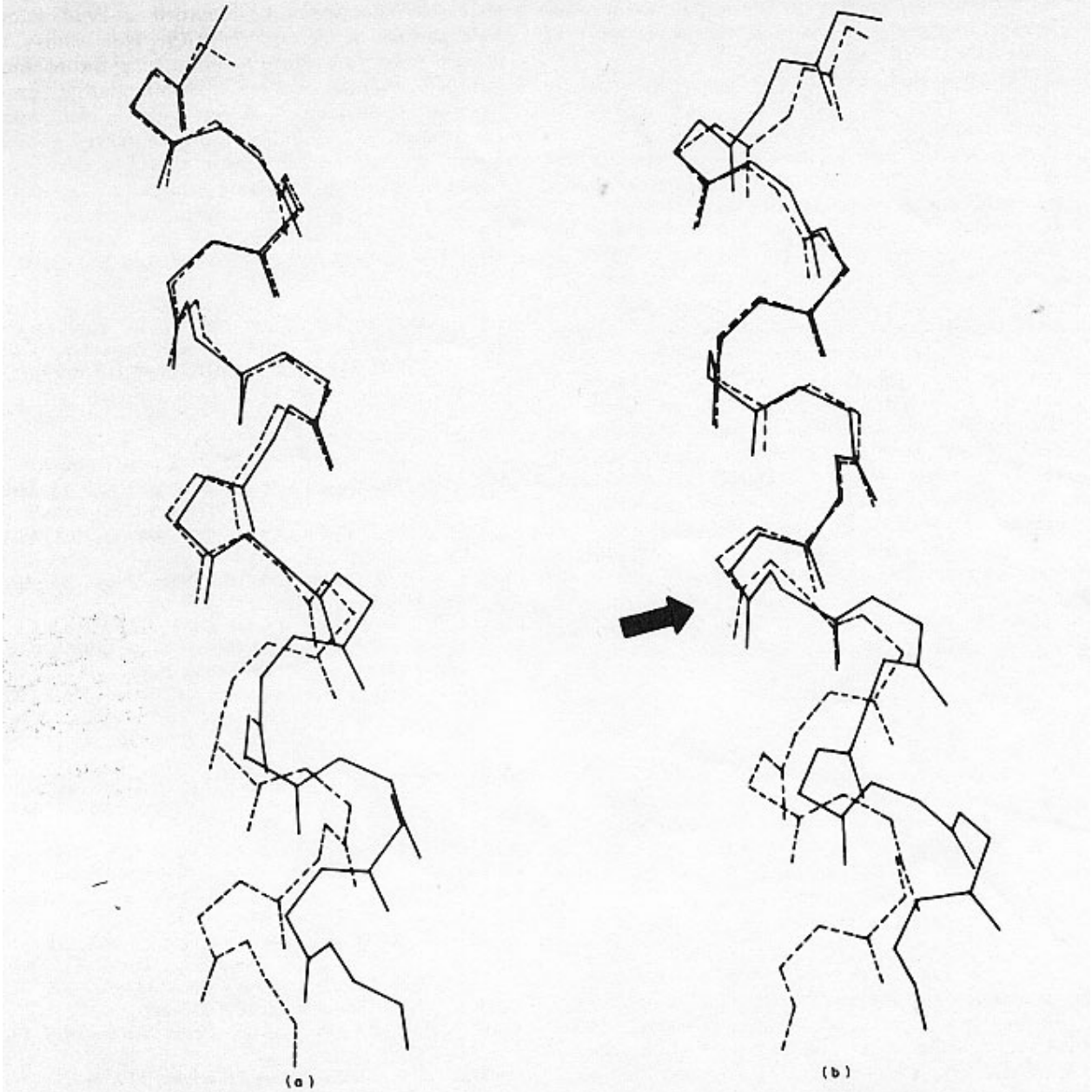

Figure 14. (a) A curved helix found in avian pancreatic polypeptide (shown by continuous lines) superposed on to a linear helix found in phospholipase (shown by broken lines). The least-squares fit is obtained for the 1st 2 turns in each helix. (b) The comparable view of a kinked helix from erythrocruorin (shown by continuous lines) superposed on to a linear helix found in phospholipase (shown by broken lines). The arrow indicates the position of the kink in the erythrocruorin helix.

way imposes its own restrictions on the helix position.

In the other kinked helices identified in this analysis, it is interesting to note that the kinks occur near residues close to the proteins' active sites. In carboxpeptidase, the kink in the helix involving residues 72 to 89, occurs near Thr75; this seems to be because the top part of the helix (residues 72 to 75) is tilted towards the enzyme's active site, in order that the side-chain of Glu72 can co-ordinate with the catalytic Zn^{2+} ion (Quiocho & Lipscomb, 1972). Likewise the kinks that are seen in the two helices from thermolysin (involving residues 136 to 151 and residues 159 to 180) occur near the residues Val139, which is one of the residues that lines the hydrophobic pocket surrounding the active site Zn iron, and Glu166, which is actually co-ordinated to the Zn ion *via* its side-chain carboxyl group (Matthews *et al.*, 1974). (An additional example that was reported recently in the literature involves the distal helix in cytochrome P450, which has a kink near the conserved Thr252, very close to the camphor binding site (Poulos *et al.*, 1985). There appears to

be no common pattern in the sequences of these helices, and we conclude that the kinks must be caused by packing effects.

The kinks that are caused by proline residues are tolerated only by long helices that are packed against several other secondary structures. Since the proline residues are in all cases conserved, the question arises as to whether the kinks in these helices serve some useful purpose. Certainly for some of the helices, there are indications that the kinks may have a definite functional role. In adenylate kinase, for example, the proline-kinked helix involving residues 142 to 168, is one of the structures that shifts as part of the conformational change that accompanies substrate binding (Sachsenheimer & Schulz, 1977). This shift is confined to the residues on the N-terminal side of the proline (those on the C-terminal side being unaffected by substrate binding), suggesting that the proline here acts to "uncouple" the two halves of the helix, so that the N-terminal end can move independently of the C-terminal end.

A more direct functional role for a proline-kinked helix is seen in catalase, where the kink caused by Pro358 in the α9 helix redirects the helix, in order to optimize the interaction between Tyr357 and the haem group (Murthy *et al.*, 1981).

We may note here also, that there are a number of conserved proline residues in the sequences of various ion-transporting integral membrane proteins, in regions which are predicted to form transmembrane helices (see, for example, Rao *et al.*, 1978; Noda *et al.*, 1983; Kubo *et al.*, 1986). It is tempting to speculate therefore, that these helices may in fact be kinked, and that the kinks are important for pore formation, or in signal transduction to the opening of the ion channels.

The distortions that occur at the end(s) of an α-helix represent a tightening or unravelling of the first or last turns. The reasons for these changes in conformation are not entirely clear. Although seven of the ten C-terminally kinked helices have carbonyl groups in the last turn that are hydrogen bonded to side-chains (usually serine or threonine), this can not be considered as the sole cause of their distortion. As noted by Baker & Hubbard (1984), main-chain/side-chain interactions of this sort are prevalent along the entire lengths of α-helices, and are not, therefore, a major destabilizing influence. It may be, however, that if the terminal residues have a low propensity for helix formation, then these interactions may have a more significant effect on the helix structure. Certainly, the terminal residues will need to balance their helical interactions against interactions with the rest of the protein and solvent, and the conformation that they adopt will reflect this compromise.

4. Conclusion

This survey has highlighted the distortions that arise in helices. It has shown that α-helices are more deformable than has previously been suggested, and that they should not be regarded as rigid, linear rods (Schulz & Schirmer, 1979). This ability to deform must contribute to optimizing the packing between helices, and to accommodating gross changes in sequence. It can have a functional importance, as found for some conserved proline kinks. This is the same view that has developed from studies of the dynamic behaviour of α-helices (see, for example, Chou, 1983), and is consistent with the observations that show that ΔG values for helix deformation are small (Parry & Suzuki, 1969).

We thank Professor Tom Blundell for kindling our interest in this subject, and for valuable discussions of the work. D.J.B. is supported by SERC grant GR/C/94605.

References

Arnott, S. & Wonacott, A. J. (1966). *J. Mol. Biol.* **21**, 371–383.

Artymiuk, P. J. & Blake, C. C. F. (1981). *J. Mol. Biol.* **152**, 737–762.

Baker, E. N. & Hubbard, R. E. (1984). *Progr. Biophys. Mol. Biol.* **44**, 97–179.

Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C. & Bonora, G. M. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 1988–1992.

Bernstein, F. C., Koetzle, T. F., Williams, E. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982). *J. Biol. Chem.* **257**, 13650–13662.

Blundell, T. L., Barlow, D. J., Borkakoti, N. & Thornton, J. M. (1983). *Nature (London)*, **306**, 281–283.

Chakrabati, P., Bernard, M. & Rees, D. C. (1986). *Biopolymers*, **25**, 1087–1093.

Chothia, C. (1984). *Annu. Rev. Biochem.* **53**, 537–572.

Chothia, C., Levitt, M. & Richardson, D. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4130–4134.

Chou, K. (1983). *Biochem. J.* **209**, 573–580.

Chou, C. P. Y. & Fasman, G. D. (1974). *Biochemistry*, **13**, 222–245.

Crick, F. H. C. (1952). *Nature (London)*, **170**, 882.

De Santis, P., Giglio, E., Liquori, A. M. & Ripamonti, A. (1965). *Nature (London)*, **206**, 456–458.

Donoghue, J. (1953). *Proc. Nat. Acad. Sci., U.S.A.* **39**, 470–478.

Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982). *Nature (London)*, **299**, 371–374.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T. & Numa, S. (1986). *Nature (London)*, **323**, 411–416.

Johnson, C. N. (1976). *Oak Ridge Nat. Lab. Rep.* **5138**, Ortep II.

Lesk, A. M. & Chothia, C. (1980). *J. Mol. Biol.* **136**, 225–270.

Levitt, M. (1978). *Biochemistry*, **17**, 4277–4285.

Love, W. E., Klock, P. A., Lattman, E. E., Padlan, E. A., Ward, K. B. Jr & Hendrickson, W. A. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 349–357.

Matthews, B. W., Weaver, L. H. & Kester, W. R. (1974). *J. Biol. Chem.* **249**, 8030–8044.
Murthy, M. R. N., Reid, III, T. J., Sicignano, A., Tanaka, N. & Rossman, M. G. (1981). *J. Mol. Biol.* **152**, 465–499.
Noda, M., Takahashi, H., Tanabe, T., Toyosata, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T. & Numa, S. (1983). *Nature (London)*, **302**, 528–532.
Parry, D. A. D. & Suzuki, R. B. (1969). *Biopolymers*, **7**, 189–197.
Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Nat. Acad. Sci., U.S.A.* **37**, 205–211.
Perutz, M. F. (1951). *Nature (London)*, **167**, 1053–1054.
Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–678.
Poulos, T. L., Finzel, B. C., Gunsalus, T. C., Wagner, G. C. & Kraut, J. (1985). *J. Biol. Chem.* **260**, 16122–16130.
Quiocho, F. A. & Lipscomb, L. W. N. (1972). *Advan. Protein Chem.* **25**, 1–78.
Rao, J. K., Hargrave, P. A. & Argos, P. (1983). *FEBS Letters*, **156**, 165–169.
Richardson, J. S. (1981). *Advan. Protein Chem.* **34**, 167–339.
Richmond, T. J. & Richards, F. M. (1978). *J. Mol. Biol.* **119**, 537–555.
Sachsenheimer, W. & Schulz, G. E. (1977). *J. Mol. Biol.* **114**, 23–36.
Sakabe, N., Sakabe, K. & Sasaki, K. (1981). In *Structural Studies on Molecules of Biological Interest*, (Dodson, G., Glusker, J. P. & Sayre, D., eds), pp. 509–526, Clarendon Press, Oxford.
Schulz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, Springer-Verlag, New York.
Steigemann, W. & Weber, E. (1979). *J. Mol. Biol.* **127**, 309–338.
Takano, T. & Dickerson, R. E. (1981). *J. Mol. Biol.* **173**, 79–94.
Terwilliger, T. C. & Eisenberg, D. (1982). *J. Biol. Chem.* **257**, 6016–6022.
Vainshtein, B. K., Melik-Adamyan, W. R., Barynin, V. V., Grebenko, A. I. & Borisov, V. V. (1986). *J. Mol. Biol.* **188**, 49–61.
Watson, H. C. (1969). *Progr. Stereochem.* **4**, 299–333.
Yang, C., Brown, J. N. & Kopple, K. D. (1969). *Int. J. Protein Pept. Res.* **14**, 12–20.

Edited by R. Huber