Journal of Molecular Graphics and Modelling 33 (2012) 52–60

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

PS – a program for the analysis of helix geometry

Brian J. Smith[a,b,c,*]

[a] Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia
[b] The Walter+Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia
[c] Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 7 October 2011
Received in revised form
24 November 2011
Accepted 24 November 2011
Available online 6 December 2011

Keywords:
Alpha helix
3_{10} helix
Pi helix
Radius of curvature
Phase yield
Pitch

ABSTRACT

The structure of helices within proteins is often distorted from the ideal linear topology. Curvature of the helix axis can be measured by determining the radius of a circle fit to the axis. Described here is a method of defining a curved path that places backbone atoms (usually Cα) equidistantly from the path. The variance in the distance of backbone atoms from the helix axis is minimised to produce the parametric equations that describe the intersection of a sphere and a plane. The geometric properties of the helix (including helix radius, radius of curvature, and pitch) can be readily obtained from these equations. The approach is applicable to any form of helix, can use any atom in the peptide to determine the axis, can be applied to any polypeptide including mixed α/β peptides, and does not rely on a regular spacing of peptide monomers in the polypeptide chain.

1. Introduction

In the prototypical polypeptide alpha helix the helix axis, defined as the path of simultaneous rotation and translation that relates each successive amino acid in the contiguous chain, is linear. In the helical segments in proteins this axis tends to be distorted by the interactions with other elements of the protein and with the solvent. The first survey of the curvature of helices [1] noted the preponderance of irregularities in the helices, and concluded that the majority of alpha helices are indeed curved.

Several approaches to describe the axis of a helix have been described previously. Early methods were restricted to linear axes, and often relied on the regularity of the spacing between successive amino acids. Strategies employed to define the axis included: (1) identification of the line of best fit though a selection of atoms [2], (2) superposition of an ideal helical motif [3], (3) fitting to a cylindrical surface [4], or (4) by consideration of the geometric relationship between successive residues in the chain [5–7]. Invariably, the radius of curvature of the helix had been obtained by translating the points traced by the axis to a plane in the coordinate frame and subsequent fitting to a circle.

Described here is an alternative approach that draws on the method of Åqvist [4] but without the restriction of linearity of the axis, and without the need to translate the axis for subsequent fitting to obtain the radius of curvature. Additionally, this approach is not restricted to α-amino acids, being equally applicable to β-amino acids and mixed α/β polypeptides, and can be applied to any helix, including 3_{10}, alpha, pi, polyproline, beta helices and DNA. The method determines the great circle arc that minimises the variance in the distance from this arc to a collection of points that defines each monomer in the helical chain.

2. Methods

The great circle arc can be described by a pair of parametric equations which define the intersection of a sphere with a plane, where the centre of the sphere lies on the plane. Thus, a sphere with its origin at the point $\boldsymbol{O}(a,b,c)$ is defined (in Cartesian space (x,y,z)) by Eq. (1).

$$(x-a)^2+(y-b)^2+(z-c)^2=R^2 \quad (1)$$

where R is the sphere radius, and the equation of a plane that includes the point (a,b,c) can be written

$$l(x-a)+m(y-b)+n(z-c)=0 \quad (2)$$

Thus, the great circle arc is defined by the coordinates of the sphere origin (a,b,c), the sphere radius, R, and the plane

* Correspondence address: Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia.
Tel.: +61 03 9479 3245; fax: +61 03 9479 1266.
E-mail address: brian.smith@latrobe.edu.au

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.11.004

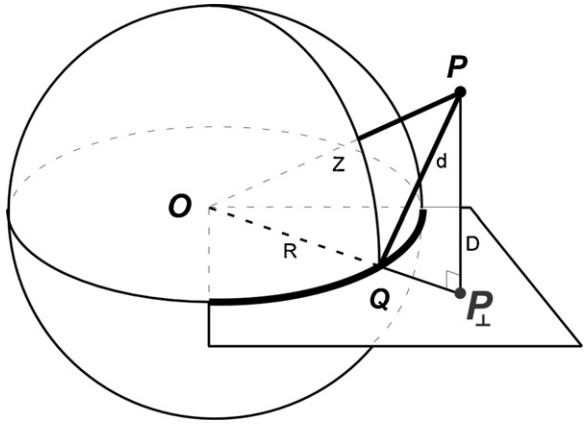

Fig. 1. Definition of the parameters used to define the helix axis. The great circle arc is defined by the intersection of a sphere of radius R centred at the origin $\boldsymbol{O}$, and a plane that includes the sphere origin. A point $\boldsymbol{P}$ that lies a distance Z from the origin is a distance D from the plane (perpendicular), and a distance d from the nearest point on the great circle arc, $\boldsymbol{Q}$. The variance in the radial distances d_k for a set of points $\boldsymbol{P_k}$ is minimised using the method of conjugate gradients to determine the coordinates of the sphere origin (a,b,c), the sphere radius (R), and the orientation of the plane (l,m,n).

perpendicular (l,m,n).The distance, Z, of the point $\boldsymbol{P}(i,j,k)$ to the sphere origin $\boldsymbol{O}$ (Fig. 1) is obtained from Eq. (3).

$$(i-a)^2+(j-b)^2+(k-c)^2=Z^2 \tag{3}$$

and the perpendicular distance, D, of the point $\boldsymbol{P}$ from the plane can be obtained from Eq. (4).

$$\frac{(l(i-a)+m(j-b)+n(k-c))^2}{(l^2+m^2+n^2)}=D^2 \tag{4}$$

The distance, d, of the point $\boldsymbol{P}$ from the great circle arc is then related to Z and D by the relationship

$$Z^2+R^2-2R(Z^2-D^2)^{1/2}=d^2 \tag{5}$$

Here, as stated earlier, the axis of a helix is defined by the path of the great circle arc that passes equidistantly from the collection of points that represent each residue in the chain. The requirement of equidistant separation can be achieved by minimising the standard statistical variance, s^2, of the collection of distances d.

$$s^2=\frac{1}{N-1}\sum_{k=1}^{N}(d_k-\bar{d})^2 \tag{6}$$

where, for N points ($\boldsymbol{P}_k$, $k=1, N$) a distance d_k from the axis, the average separation is given by

$$\bar{d}=\frac{1}{N}\sum_{k=1}^{N}d_k \tag{7}$$

Here, the variance is minimised by the method of conjugate gradients that utilises the following derivative of s^2 with respect to any variable $\zeta=(R,a,b,c,l,m,n)$,

$$\frac{d(s^2)}{d\zeta}=\frac{1}{N-1}\sum_{k=1}^{N}\left(1-\frac{\bar{d}}{d_k}\right)\frac{d(d_k^2)}{d\zeta} \tag{8}$$

The derivatives of d^2 with respect to each of the variables that define the great circle arc are as follows:

$$\frac{d(d^2)}{dR}=2R-2(Z^2-D^2)^{1/2} \tag{9}$$

$$\frac{d(d^2)}{da}=2(a-i)-\frac{2R(a-i-l\vartheta)}{(Z^2-D^2)^{1/2}} \tag{10}$$

$$\frac{d(d^2)}{dl}=\frac{2R\vartheta(a-i-l\vartheta)}{(Z^2-D^2)^{1/2}} \tag{11}$$

where

$$\vartheta=\frac{(l(a-i)+m(b-j)+n(c-k))}{l^2+m^2+n^2} \tag{12}$$

and, where the derivative with respect to b and c are available by the apparent substitutions in Eq. (10), and the derivatives with respect to m and n are available by the apparent substitutions in Eq. (11).

The coordinates of the point on the plane perpendicular to $\boldsymbol{P}$, $\boldsymbol{P}_\perp(i_\perp,j_\perp,k_\perp)$, are

$$i_\perp=i-l\vartheta \tag{13}$$

and, where the coordinates $j_\perp$ and $k_\perp$ are available by apparent substitutions in Eq. (13). The coordinates, $\boldsymbol{Q}(u,v,w)$, of the point on the great circle arc perpendicular to $\boldsymbol{P}$ are then given by Eq. (14).

$$u=a+\frac{R(i_\perp-a)}{((i_\perp-a)^2+(j_\perp-b)^2+(k_\perp-c)^2)^{1/2}} \tag{14}$$

and, where the coordinates v and w are available by apparent substitutions in this equation. Eq. (14) can also be used to determine the distance of any point from the axis.

For linear helices the rise of the helix from one residue to the next, τ, is simply the distance between successive points $\boldsymbol{Q}$ (where all the points $\boldsymbol{Q}$ correspond to the same atom type, for example, Cα atoms). For curved helices, however, the rise is obtained from the arc length

$$\tau_k=2R\,\sin^{-1}\left(\frac{Q_kQ_{k-1}}{2R}\right) \tag{15}$$

The helix length, L, rather than being the Cartesian distance between the first and last $\boldsymbol{Q}$s along the helix axis, is given by the arc length between the first point $\boldsymbol{Q}_1$ and the last point $\boldsymbol{Q}_N$ in Eq. (15).

The phase yield (ϕ, the rotation about the helix axis from one residue to the next) of a linear helix is obtained from the dot product of the vectors $\boldsymbol{P}_{k-1}\boldsymbol{Q}_{k-1}$ and $\boldsymbol{P}_k\boldsymbol{Q}_k$. For curved helices the vector $\boldsymbol{P}_k\boldsymbol{Q}_k$ must first be projected onto a linear axis formed by the line at a tangent to the axis at the origin of the vector $\boldsymbol{P}_{k-1}\boldsymbol{Q}_{k-1}$. The transformation requires rotation of the vector $\boldsymbol{P}_k\boldsymbol{Q}_k$ by an angle $-\tau_k/R$ about an axis perpendicular to the plane of the great circle and centred at the vector origin ($\boldsymbol{Q}_k$), followed by a translation to the tangential line.

The average number of residues per turn of the helix, T_{av}, is obtained from Eq. (16).

$$T_{av}=\frac{2\pi}{\phi_{av}} \tag{16}$$

where ϕ_{av} is the average phase yield. The average pitch, the vertical distance between successive turns of the helix, p_{av}, is then given by

$$p_{av}=\frac{2\pi L}{(N-1)\phi_{av}} \tag{17}$$

The relationship between the helix parameters p pitch, d helix, radius, τ rise, and ϕ phase yield, is illustrated in Fig. 2.

The program PS (Pandus Semita, *Lat.*, curved path) determines the parameters that define the great circle arc, namely a, b, c, l, m, n and R by the method of conjugate gradient minimisation of the variance, s^2, of the radial distances d_k to the great circle arc.

The method of conjugate gradients cannot guarantee to locate the global minimum; applying different initial conditions permits identification of the appropriate great circle arc parameters. Since 7 parameters are required to describe the great circle arc, at least 7 points are needed to perform the helix analysis. Helices that are (near) exactly linear will have an extremely large radius of curvature; in these circumstances, refinement of the parameters that define the great circle arc becomes hypersensitive to the

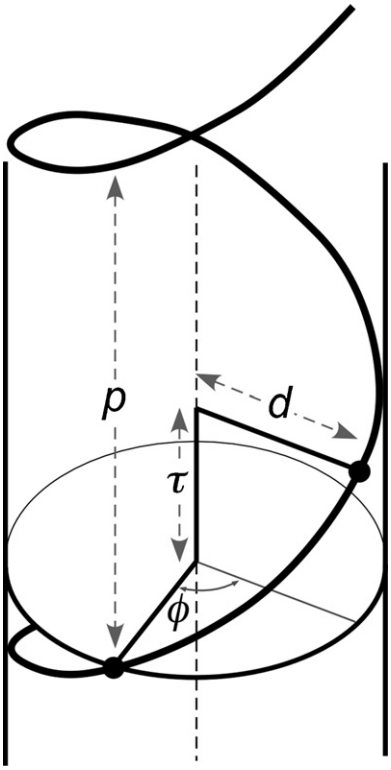

Fig. 2. Diagram illustrating the helix parameters, p – pitch; d – helix radius; τ – rise; ϕ – phase yield.

coordinates of the atoms used to define the helix and to the initial conditions used to begin the conjugate gradient minimisation (small changes in coordinate result in seemingly large changes in parameters). However, the geometric properties of the helix (radius, phase yield, pitch, etc.) are (essentially) unaffected – they remain equivalent within the limits of the observed variation. This property of the method is not limiting, since helices in proteins are seldom (if ever) exactly linear. The source code for the program, written in FORTRAN, is available in the Supporting Information.

3. Results and discussion

The 3_{10}, alpha and pi helices are characterised by inter-residue hydrogen bonding between the backbone carbonyl oxygen of one residue with the backbone amide hydrogen of another, where the number of intervening residues is 2, 3 or 4, respectively. The polyproline helices have no internal hydrogen bonding. In the type-II polyproline helices the amide bond adopts the *trans* configuration, resulting in a highly extended conformation, whereas a *cis* configuration of the amide bond in type-I polyproline helices results in a significantly more compact conformation.

The PS analysis of helices can be performed on any atoms in the helix. Thus, whereas the Cα carbon has traditionally been used to characterise the helix, the PS analysis can use, for example, the carbonyl carbon or oxygen, or the amide nitrogen of the polypeptide backbone, or the Cα carbon. In Table 1 is presented an analysis of the five different helix types using either the carbonyl (C) or Cα carbon, or amide nitrogen (N) atoms to define the helix axis (ideal 20-residue helix coordinates are provided in the Supporting Information). Other than the helix radius, which clearly depends on the atom used to define the axis, all other parameters defining the helix (rise per residue, phase yield, residues per turn, and pitch) are independent of atom type. The helix radius of the type-II polyproline helix using the carbonyl carbon atom to define the axis is of particular note – this atom lies close to the helix axis, resulting in a helix radius of 0.32 ± 0.01 Å.

The PS analysis generally yields helix parameters similar to those obtained from other methods, in particular the HBEND [3], HELANAL [5] and TWISTER [7] programs.

In the survey by Barlow and Thornton [3] 15 proteins containing a total of 45 helices were examined using the program HBEND. A comparison of the results from PS with those from HBEND (Table 2) shows a generally good correlation of the calculated radius of curvature for linear and curved helices, although for the linear helices the radii from PS are generally larger than those reported using HBEND. For the irregular and kinked helices the PS and HBEND radii are either in excellent agreement, or the PS radii are significantly larger than those reported using HBEND. Also, for these helices the calculated standard deviation of the helix radius from PS tends to be large, >0.1, and the range of radial distances (from smallest to largest) also tends to be large, up to 0.7 Å. Thus, kinked and irregular helices can generally be identified by large deviations in the radial distances using PS.

The structure of carboxypepdidase (5CPA) contains alpha helices of each different classification, curved, irregular and kinked. The helix in the residue range 72–89 was classified as kinked in the analysis using HBEND. The radius of curvature calculated using PS and HBEND are identical, 33 and 34 Å, respectively (Fig. 3). The radial distances calculated using PS vary between 2.08 and 2.49 Å across this residue range, whereas if residues before the kink (72–76) are not included in the analysis, the radius of curvature, R, reduces slightly to 27 Å, and the variation in radial distances reduces significantly (2.17–2.40 Å). The deviation from the ideal alpha helix radius for each residue is presented in Fig. 3 (for this helix and others). The helix in the residue range 216–231 was classified as linear in the HBEND analysis. The PS analysis yields a slightly larger radius of curvature than HBEND (130 compared with 111 Å), and produces radial distances that deviate little from the ideal helix radius. The residue range 286–306 was classified as irregular from the HBEND analysis. In the PS analysis the average helix radius, 2.31 ± 0.17 Å, is associated with a very large standard deviation, due to the large variation in radial distances measured (2.02–2.63 Å).

While the PS and HBEND analyses generally provide comparable results, three differences are worthy of note. In lysozyme (4LZT), the helix in the residue range 6–15 is characterised as irregular in the HBEND analysis [3], and the radius of curvature is particularly small (26 Å). Across this same residue range the PS analysis yields a significantly larger radius of curvature, 105 Å, with a helix radius of 2.28 ± 0.04 Å. The average rise per residue, 1.50 ± 0.10 Å, average phase yield, $99.9 \pm 2.3^{\circ}$, and number of residues per turn, 3.60 ± 0.09, are typical of an alpha helix. In myoglobin (1MBD), the helix in the residue range 20–36 is also characterised as irregular using the HBEND analysis, the radius of curvature is small (41 Å) and the axis is poorly defined [3]. In contrast the radius of curvature from the PS analysis is large (345 Å), consistent with a linear helix, and the helix is reasonably well defined (standard deviation in the helix radius of 0.09 Å). Similarly, in dihydrofolatereductase (3DFR) the helix defined by residues 78–87 has a very small radius of curvature based on the HBEND analysis (21 Å), yet the PS analysis yields a radius of curvature significantly larger (71 Å). Since the PS method minimises the variation in the radial distance d, the average helix radius obtained from PS shows less variation (smaller standard deviation) than if the radius of curvature obtained from HBEND is used to define the helix axis. In the three examples above, if the radius of curvature from the HBEND analysis is used to determine the helix axis, the helix radius obtained from the PS analysis is 2.22 ± 0.20, 2.29 ± 0.35 and 2.25 ± 0.26 Å, respectively, exhibiting significantly larger deviation in the radial distances than those obtained when the radius if curvature is determined using PS.

Table 1
Comparison of helix parameters for ideal helices.[a]

Helix[b]	Helix radius[c] (Å)	Rise per residue (Å)	Phase yield (°)	Residues per turn	Pitch (Å)
3_{10} (Cα)	1.92(01)	1.91(01)	119.0(0.2)	3.02(01)	5.76(01)
3_{10} (C)	1.27(01)	1.91(01)	119.0(0.9)	3.02(02)	5.76(04)
3_{10} (N)	1.10(01)	1.91(01)	119.1(0.6)	3.02(02)	5.76(03)
Alpha (Cα)	2.30(01)	1.48(01)	99.9(0.1)	3.60(01)	5.34(01)
Alpha (C)	1.68(01)	1.48(01)	99.8(0.1)	3.61(01)	5.34(01)
Alpha (N)	1.56(01)	1.48(01)	99.9(0.2)	3.61(01)	5.34(01)
Pi (Cα)	2.68(01)	1.02(01)	86.8(0.2)	4.15(01)	4.22(01)
Pi (C)	2.01(01)	1.02(01)	86.7(0.4)	4.15(02)	4.22(02)
Pi (N)	2.06(01)	1.02(01)	86.8(0.2)	4.15(01)	4.22(01)
Polyproline-I(Cα)	1.27(01)	1.83(01)	116.3(0.2)	3.10(01)	5.67(01)
Polyproline-I(C)	1.52(01)	1.83(01)	116.3(0.2)	3.10(01)	5.67(01)
Polyproline-I(N)	1.86(01)	1.83(01)	116.3(0.1)	3.10(01)	5.67(01)
Polyproline-II(Cα)	1.30(01)	3.12(01)	118.8(0.4)	3.03(01)	9.44(03)
Polyproline-II(C)	0.32(01)	3.11(01)	119.0(2.8)	3.02(07)	9.42(22)
Polyproline-II(N)	1.06(02)	3.12(01)	118.9(1.4)	3.03(04)	9.43(11)

[a] Determined using fine-grain scan of initial seed for conjugate gradients optimization.
[b] Atoms used to define the helix are in parenthesis.
[c] Numbers in parenthesis are single standard deviation.

The PS and HELANAL [5] programs yield comparable analyses of the radius of curvature (Table 3), especially for the curved helix in G-CSF (1BGE). Although the radius of curvature for the helix spanned by residues 102–129 in the ribonucleotidereductase protein R2 (1RIB) using PS (306 Å) is significantly larger than that reported using HELANAL (140 Å), it is clear both approaches indicate a near linear helix. The helix spanning residues 44–74 in the structure of lysin (1LIS) contains a kink near residue 61. The helix radius produced from PS is significantly larger than the ideal alpha helix radius and exhibits a very large variation (2.55 ± 0.80 Å); such poor fit is diagnostic of a significant deviation from a simple curved or linear topology. However, the N-terminal and C-terminal helices either side of residue 61 can be well fit to two individual helices, each exhibiting a highly curved topology with a radius of curvature of 40 and 30 Å, respectively. The quantitative analysis of helices, including helix radius and radius of curvature, using the PS approach is predicated on the assumption that the helix axis can be described by a simple great circle arc. As shown above, this can be achieved in kinked helices by subdividing the helix into multiple components.

The PS approach is not particularly well suited to the analysis of long coiled-coils due to the super-helix winding of the

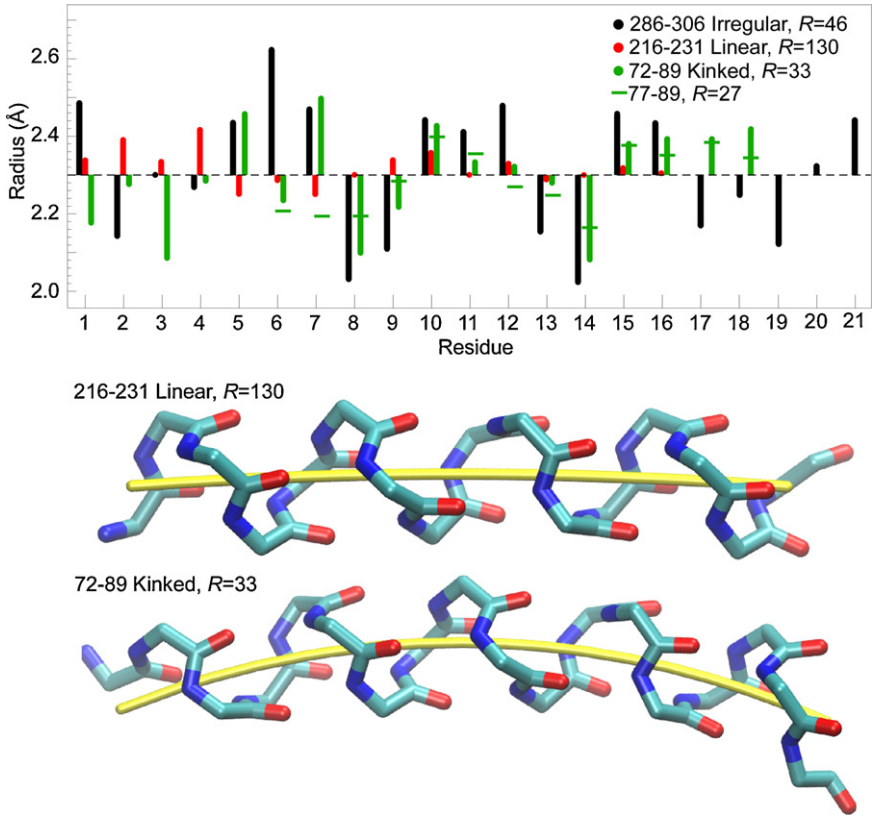

Fig. 3. Comparison of helix Geometry. Presented in the top panel is the deviation of radial distances from the ideal α-helix radius (of 2.30 Å) for three helices in carboxypeptidase (5CPA). Green horizontal lines correspond to the radial distances calculated when the first 5 residues of the kinked helix (72–76) are removed from the analysis. In the bottom two panels the trace of the helix is shown (in yellow) against the backbone atoms (carbon, cyan; nitrogen, blue; oxygen, red) of two helices, the linear helix spanning residues 216–231, and the kinked helix spanning residues 72–89. The radius of curvature in these two helices is 130 and 33 Å, respectively.

Table 2
Comparison of helix radius and radius of curvature from PS and HBEND.

Protein (PDB)		Chain	Residue range	Helix radius[b] (Å)	Min. radial distance (Å)	Max. radial distance (Å)	Radius of curvature (Å)	HBEND[a]	
								Radius (Å)	Desc.[c]
4LZT[d]	Lysozyme	A	6–15	2.28(04)	2.22	2.33	105	26	Irregular
			25–35	2.31(06)	2.21	2.42	58	58	Curved
			89–99	2.28(07)	2.15	2.40	90	57	Curved
7RSA[e]	Ribonuclease	A	3–12	2.32(08)	2.16	2.46	37	29	Curved
			24–33	2.33(09)	2.19	2.48	46	34	Curved
2P8O[f]	Chymotrypsin	C	236–243	2.30(01)	2.29	2.32	150	62	Curved
5CPA[g]	Carboxypeptidase	A	15–27	2.31(06)	2.15	2.40	69	62	Curved
			72–89	2.30(13)	2.08	2.49	33	34	Kinked
			173–187	2.33(06)	2.24	2.46	132	71	Irregular
			216–231	2.32(04)	2.25	2.42	130	111	Linear
			286–306	2.31(17)	2.02	2.63	46	45	Irregular
2TMN[h]	Thermolysin	E	67–87	2.38(16)	1.99	2.63	145	89	Kinked
			136–151	2.32(10)	2.07	2.51	36	35	Kinked
			160–179	2.31(10)	2.16	2.49	206	127	Kinked
			234–246	2.35(11)	2.18	2.58	77	78	Curved
			261–272	2.34(07)	2.23	2.43	82	71	Kinked
			281–295	2.30(08)	2.19	2.44	234	100	Linear
			300–313	2.27(12)	2.09	2.52	57	69	Curved
1MBD[i]	Oxymyoglobin	A	4–18	2.30(10)	2.09	2.49	48	49	Curved
			20–36	2.36(09)	2.20	2.50	345	41	Irregular
			59–77	2.30(08)	2.11	2.39	83	85	Curved
			82–94	2.28(12)	2.11	2.54	26	24	Kinked
			101–118	2.29(07)	2.14	2.40	208	184	Linear
			124–149	2.30(12)	2.11	2.56	92	88	Curved
9PAP[j]	Actinidin	A	25–42	2.35(13)	2.07	2.62	87	78	Curved
			69–78	2.32(08)	2.17	2.45	45	31	Curved
			120–128	2.28(03)	2.24	2.32	29	34	Curved
1CRN[k]	Crambin	A	6–17	2.30(05)	2.23	2.39	89	60	Curved
1PPT[l]	aPP	A	15–32	2.32(05)	2.24	2.39	65	68	Curved
5CYT[m]	Cytochrome *c*	R	2–13	2.31(07)	2.20	2.43	47	48	Curved
			87–100	2.28(04)	2.22	2.34	91	94	Linear
2OVO[n]	Ovomucoid	A	34–43	2.32(09)	2.18	2.45	59	54	Curved
2MLT[o]	Melittin	A	1–9	2.28(02)	2.27	2.31	23	63	Curved
			13–26	2.28(06)	2.18	2.38	86	89	Curved
1ECA[p]	Erythrocruorin	A	2–15	2.27(10)	2.09	2.41	55	72	Curved
			20–31	2.30(07)	2.18	2.41	67	63	Curved
			52–72	2.28(15)	2.04	2.51	45	43	Kinked
			76–86	2.29(06)	2.21	2.40	76	80	Curved
			93–110	2.27(07)	2.09	2.37	129	112	Linear
			117–134	2.34(14)	2.06	2.57	60	50	Irregular
1BP2[q]	Phospholipase	A	3–11	2.29(02)	2.27	2.31	54	53	Curved
			39–54	2.28(09)	2.12	2.40	144	112	Linear
			89–106	2.34(11)	2.05	2.50	71	71	Curved
3DFR[r]	Dihydrofolate reductase	A	23–3178–87	2.29(04)2.31(07)	2.232.20	2.352.46	4671	4621	CurvedIrregular

[a] Ref. [3].
[b] Numbers in parenthesis are single standard deviation.
[c] Qualitative description of helix geometry from an analysis using HBEND.
[d] Ref. [8].
[e] Ref. [9].
[f] Ref. [10].
[g] Ref. [11].
[h] Ref. [12].
[i] Ref. [13].
[j] Ref. [14].
[k] Ref. [15].
[l] Ref. [16].
[m] Ref. [17].
[n] Ref. [18].
[o] Ref. [19].
[p] Ref. [20].
[q] Ref. [21].
[r] Ref. [22].

helices, whereas the TWISTER program [7] is specifically designed to analyse these structures. Both the PS and TWISTER programs can accommodate β-amino acids in the polypeptide chain – in β-amino acids the Cβ carbon atom forms part of the polypeptide backbone between the amino and Cα atoms. In the PS analysis an atom other than the Cα can be used to map the helix axis to limit any bias in the resulting axis of the β-carbon: In the analysis reported in Table 4 we use the carbonyl carbon atom (C) to determine the great circle parameters of mixed α/β-polypeptides.

The leucine zipper GCN4 is a short (31–residue) coiled-coil that is capable of adopting a variety of different topologies (trimer, 4-helix parallel bundle, 4-helix anti-parallel bundle) depending on the amino acid composition. Presented in Table 4 are the results of an analysis of three GCN4 leucine zipper molecules, and the

Table 3
Comparison of helix radius and radius of curvature from PS and HELANAL.

Protein (PDB)	Residue range	Helix radius (Å)	Min. radius (Å)	Max. radius (Å)	Radius of curvature (Å)	HELANAL[a]	
						Radius	Desc.[b]
1RIB[c]	102–129	2.29(15)	2.08	2.60	306	140	Linear
1BGE[d]	144–169	2.32(16)	1.90	2.57	73	73	Curved
1LIS[e]	44–74	2.55(80)	1.43	3.86	31	25	Kinked
1LIS	44–61	2.31(09)	2.12	2.46	40		
1LIS	62–74	2.30(08)	2.21	2.45	30		
1ARS[f]	397–405	2.32(10)	2.17	2.46	63	170	Curved
2ZTA(A)[g,h]	1–30	2.28(13)	2.02	2.55	146	141	Curved
2ZTA(B)[g,i]	1–30	2.31(09)	2.09	2.47	109	107	Curved

[a] Ref. [5].
[b] Qualitative description of helix geometry from an analysis using HELANAL reported in Ref. [5].
[c] Ref. [23].
[d] Ref. [24].
[e] Ref. [25].
[f] Ref. [26].
[g] Ref. [27].
[h] A chain.
[i] B chain.

corresponding α/β-polypeptide analogues of each. The helix radius (to the Cα atom) calculated using PS is in close agreement with that reported [28] using the TWISTER program for both the coiled-coiled trimer (1IJ2) and 4-helix parallel bundle (1GCL).

There is also excellent concordance between the two programs for the helix radius for each of the α/β-polypeptide analogues. The helix radius of the α/β-polypeptide analogues is ∼0.18 Å larger than the all-α peptide, whether in the GCN4 leucine zipper or the HIV gp41 6-helix bundle. The PS calculated standard deviation in the helix radius in these coiled-coil structures tends to be larger than that calculated in helices embedded in globular proteins. The α/β-polypeptides also tend to have a larger radius of curvature compared with their all-α counterparts.

The phase yield, the rotation around the helix axis from one amino acid to the next in the polypeptide chain, is 100° in an ideal linear alpha helix. In the all-α amino acid coiled-coils listed in Table 4 the phase yield is ∼100°. In the α/β-polypeptides the phase yield produced by the β-amino acids, ∼122°, is significantly larger than the phase yield produced by the α-amino acids, ∼93°, in order to accommodate the additional methylene carbon in the polypeptide backbone. The α/β-polypeptides listed in Table 4 fall into one of three categories, (1) two amino acids in a heptad are replaced with β–amino acids, (2) an *aaab* sequence repeat, or (3) an *aab* sequence repeat of amino acids. The phase yield of one α/β-heptad is ∼708°, roughly equivalent to the phase yield in an all-α amino heptad. The phase yield of an *aaab* or an *aab* sequence is 401° or 308°, respectively, roughly equivalent to that expected for a quartet or triplet of all-α amino acids, respectively. This equivalence in phase yield within each repeat allows the mixed α/β-polypeptides to mimic the packing observed in the helical bundles of the all-α amino acids.

Table 4
Analysis of the helix radius of coiled-coils and α/β-polypeptides.

PDB		Residue range	Helix radius (C)[a,b] (Å)	Radius of curvature (Å)	Phase yield[b,c,d] (°)	Helix radius (Cα)[a,b,d] (Å)	TWISTER[e]
GCN4 (leucine zipper coiled-coil trimer)							
1IJ2[f]	α	1–29	1.71(08)	112	99.3(3.4)	2.29(08)	2.30[d]
2OXJ[g]	α/β	2–30	1.89(14)	386	91.8(4.0)/124.4(6.8)	2.45(16)	2.44[d]
GCN4 (leucine zipper 4-helix parallel bundle)							
1GCL[h]	α	3–29	1.73(19)	139	100.1(6.7)	2.27(18)	2.25[g]
2OXK[f]	α/β	1–30	1.87(09)	153	93.1(3.3)/122.6(4.3)	2.44(14)	2.42[g]
3C3F[i]	α/β	1–24	1.84(16)	109	95.8(5.5)/121.3(6.0)	2.43(18)	2.39[h]
3C3G[i]	α/β	2–28	1.91(14)	158	93.2(4.0)/116.2(5.8)	2.50(21)	2.44[h]
3C3H[i]	α/β	1–29	1.90(15)	190	94.0(3.6)/120.2(5.1)	2.48(23)	2.43[h]
GCN4 (leucine zipper 4-helix anti-parallel bundle)							
2CCF[j]	α	1–30	1.68(12)	149	101.0(4.7)	2.32(13)	
3F86[k]	α/β	2–30	1.89(06)	198	91.8(2.1)/123.7(1.9)	2.48(09)	
3F87[k]	α/β	2–31	1.90(07)	162	91.6(2.7)/124.0(3.8)	2.47(09)	
gp41 (6-helix bundle)							
3F4Y[l]	α	1–34	1.73(17)	159	100.1(6.4)	2.30(17)	
3F4Z[l]	α/β	2–37	1.91(18)	311	92.9(6.4)/123.8(6.2)	2.47(19)	

[a] Atoms used to measure the helix radius are given in parenthesis. The carbonyl carbon atoms (C) were used to determine the axis parameters.
[b] Numbers in parenthesis are single standard deviation.
[c] Average phase yield produced by α-residues (/and β-residues).
[d] Average helix parameters to Cα atoms.
[e] Helix radius (to Cα) from TWISTER.
[f] Ref. [29].
[g] Ref. [28].
[h] Ref. [30].
[i] Ref. [31].
[j] Ref. [32].
[k] Ref. [33].
[l] Ref. [34].

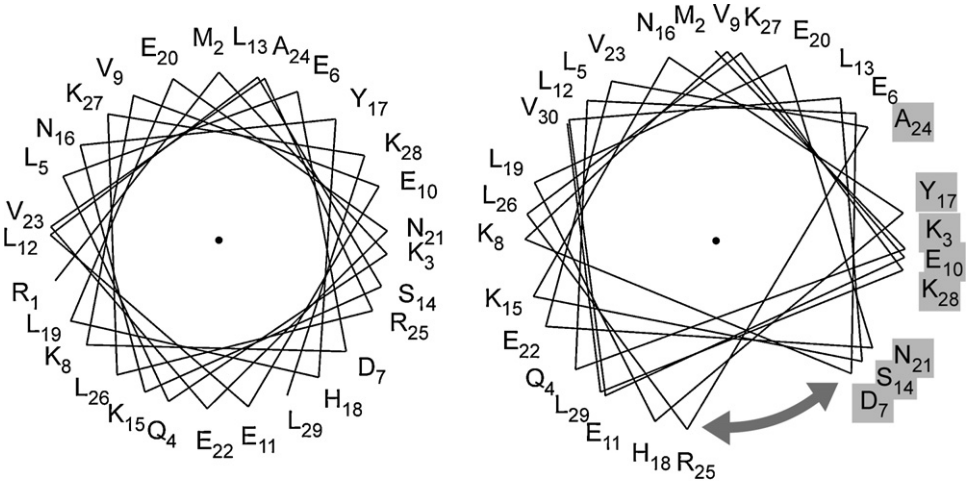

Fig. 4. Helical-wheel representation of theleucine-zipper coiled-coil trimers. In the left-hand panel the all-α amino acid peptide (1IJ2), and in the right-hand panel the α/β-polypeptide equivalent (2OXJ). The β-amino acid residues are shaded in grey. The average helix radius of the α/β-polypeptide (1.89 Å) is larger than the all-α polypeptide (1.71 Å). The double-ended arrow highlights the gap introduced into the helical wheel by the methylene carbons of the β-residues D_7, S_{14} and N_{21}. The methylene carbon atoms of K_3, E_{10}, Y_{17} and K_{28} coincide with the Cα carbon positions of D_7, S_{14} and N_{21}.

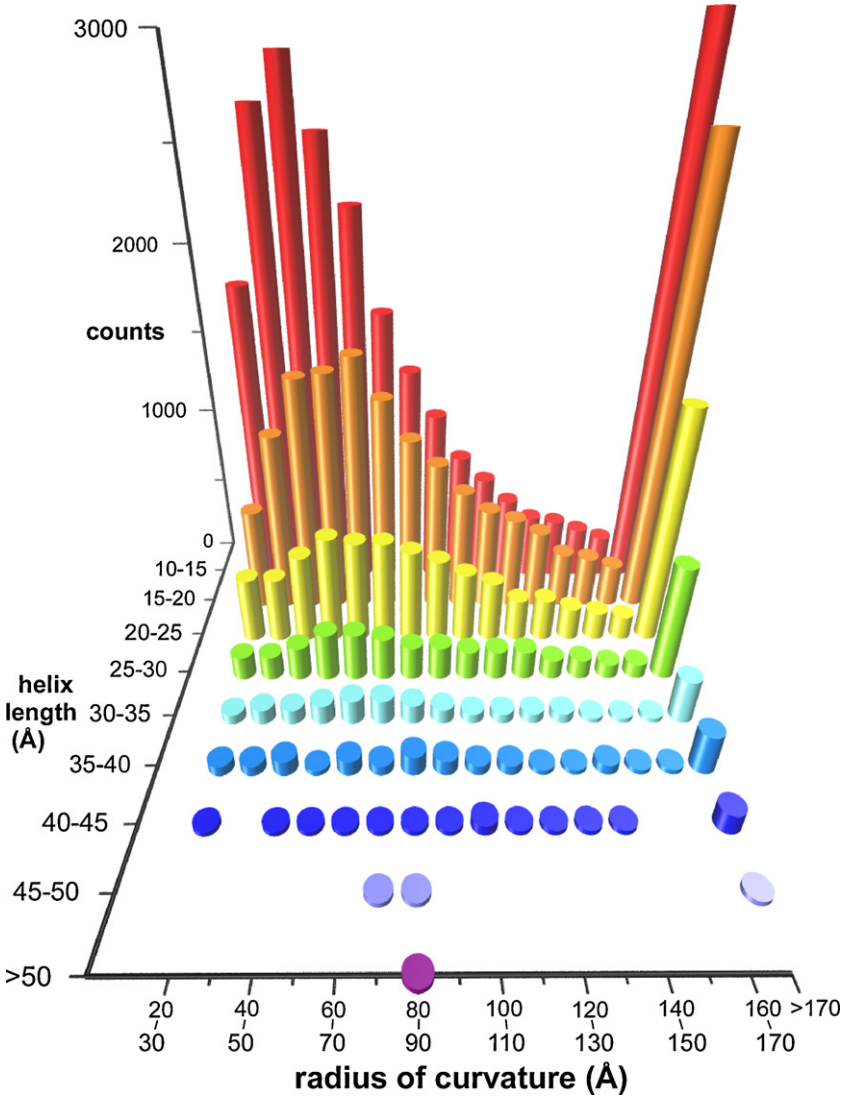

Fig. 5. Analysis of alpha helices in high-resolution X-ray protein structures. The distribution of heliceswith respect to helix length and radius of curvature.

The phase yield can be used to construct a helical wheel representation of the peptide. In Fig. 4 is presented the helical wheel representation of two leucine-zipper coiled-coil trimer helices, an all-α amino acid polypeptide (1IJ2) and a sequence-equivalent mixed α/β-polypeptide (2OXJ). The two peptides form a near-linear alpha helix (Table 4). In the all-α amino acid polypeptide the Cα atoms are (approximately) equally distributed about the helix, whereas in the mixed α/β-polypeptides the distribution of Cα atoms is less regular. The methylene carbons in the β-peptides introduce a gap into the helical wheel–this is particularly evident by the gap following the β-residues D_7, S_{14} and N_{21}. The gap following the β-residues K_3, E_{10}, Y_{17} and K_{28} coincide with the Cα carbon positions of D_7, S_{14} and N_{21}.

In their analysis of alpha helices in 205 non-homologous proteins whose structure had been determined to high resolution (<2.5 Å) Kumar and Bansal [35] determined the mean radius of curvature in curved helices was 65 ± 33 Å, although long helices were observed to be generally less curved than shorter helices. In Fig. 5 is presented an analysis of 64,536 alpha helices (with 10 or more residues) taken from 8121 protein X-ray crystal structures obtained from the PDB [36] with a resolution of 1.8 Å or better, and with sequence identity of less than 90%. Helices with a radius of curvature of less than 10 Å were clearly linked, while those with radius of curvature less than 20 Å were also kinked or irregular – the irregularity was often associated with residues at the termini of the helix. From Fig. 5 it is clear that shorter helices tend to have a smaller radius of curvature, and are therefore more curved – the shortest helices (10–15 Å) have a radius of curvature with mode of 40–50 Å, whereas helices of 30–35 Å length have a radius of curvature with mode of 70–80 Å. Helices in the 35–40 Å length range have a median radius of curvature of 100 Å, and are thus equally distributed between curved and linear.

In the first extensive survey of the geometry of helices in proteins [3] a set of criteria were developed to distinguish between (near) linear, curved, kinked and irregular alpha helices. Linear helices were defined as those whose radius of curvature is >90 Å, whereas curved helices were those <90 Å. In a more recent and more extensive analysis [35] the distinction between linear and curved geometry was based on the relative goodness of fit to either a straight or curved line, respectively. The analysis presented in Fig. 5 illustrates how the radius if curvature of helices in proteins follows a Boltzmann-like distribution in which there is no distinction between different classifications of helices. Thus, any distinction in helix geometry, linear or bent, is arbitrary, and while those criteria developed previously can be applied to any analysis from PS it is preferable to report the radius of curvature when describing the extent to which the helix is bent.

The PS analysis can equally be applied to beta helices. For example, the beta-helix domain of the carboxysomalγ-carbonic anhydrase [37] (3KWE) is shown in Fig. 6. The helix axis is defined by the Cα atoms of the hydrophobic residues that form the core of the beta helix. The helix axis is (essentially) linear, with a large radius of curvature, 214 Å – the helix radius is 5.28 ± 0.34 Å. The average rise per residue and pitch is 1.58 ± 0.63 and 4.80 ± 0.41 Å, respectively. The pitch corresponds to the average separation between β-strands in each β-sheet, and is equivalent to the rise accompanying three β-strands. However, the beta helix is better described as being assembled of layers of beta rings, where three β-strands form a single ring. The average rise per beta strand in a ring is 1.20 ± 0.24 Å, and the average rise per ring in the helix is 2.46 ± 0.14 Å in this example.

The PS method can be also applied to the helices in DNA and RNA, right or left-handed helices, polysaccharide helices, or triple helices (see electronic supplementary information).

The approach described here for determining the parametric equations of a curved helix axis yields geometrical parameters (helix radius, radius of curvature) of helices similar to those from methods reported previously. However, this approach has many advantages over the earlier methods, it can be applied to any type of helix (including beta helices, DNA and polysaccharides), it can use any atom of the polymer backbone to define the axis, it does not rely on regularly spaced monomers – indeed, it is capable of accommodating occasional missing monomers (with little bias on the resulting axis) – and, geometrical parameters describing the helix (such as helix radius and pitch) and the distance of any atom in the chain to the axis can be calculated readily.

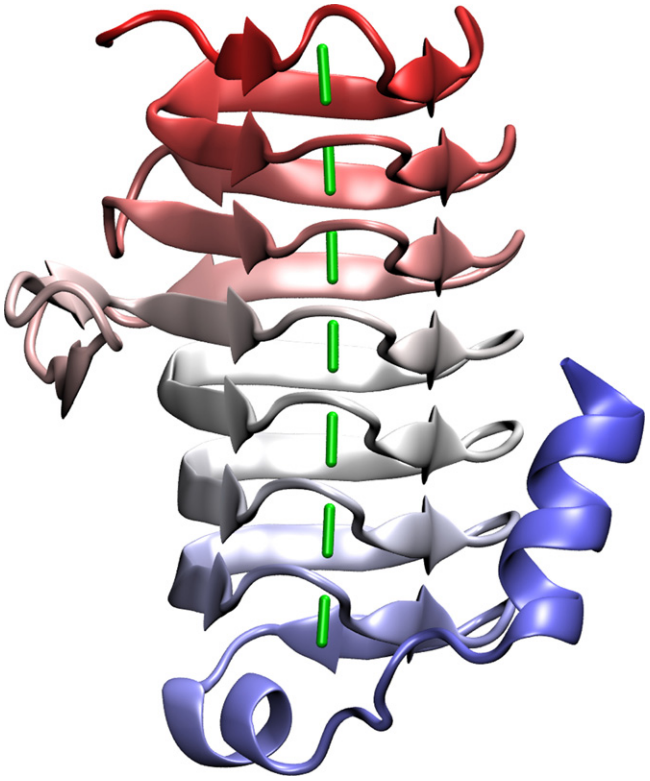

Fig. 6. Helix axis in a beta-helix. Cartoon diagram of the beta-helix domain of the carboxysomal γ-carbonic anhydrase – colouring is graded from red at the N-terminus, through white, to blue at the C-terminus. The beta-helix axis, shown in green, is defined by the Cα atoms of the central hydrophobic residues that pack in the central core of the helix. This axis is separate into the individual rings (or layers) that form the intact beta helix.

Acknowledgments

I wish to express my gratitude to Prof. Michael Lawrence, and Drs. Doug Fairlie and Erinna Lee for helpful discussions, and Prof. David Fairlie for coordinates of the right- and left-handed peptides. This work was supported by grants from the Australian Research Council (DP1093909) and the Leukemia and Lymphoma Society (SCOR 7015-02). Infrastructure support from National Health and Medical Research Council IRIISS grant #361646 and the Victorian State Government OIS grant is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.11.004.

References

[1] T. Blundell, D. Barlow, N. Borkakoti, J. Thornton, Solvent-induced distortions and the curvature of α-helices, Nature 306 (1983) 281–283.

[2] J.A. Christopher, R. Swanson, T.O. Baldwin, Algorithms for finding the axis of a helix: Fast rotational and parametric least-squares methods, Comput. Chem. 20 (1996) 339–345.
[3] D.J. Barlow, J.M. Thornton, Helix geometry in proteins, J. Mol. Biol. (1988) 601–619.
[4] J. Åqvist, A simple way to calculate the axis of an α-helix, Comput. Chem. 10 (1986) 97–99.
[5] M. Bansal, S. Kumar, R. Velavan, HELANAL: a program to characterize helix geometry in proteins, J. Biomol. Struct. Dyn. 17 (2000) 811–819.
[6] P.C. Kahn, Defining the axis of a helix, Comput. Chem. 13 (1989) 185–189.
[7] S.V. Strelkov, P. Burkhard, Analysis of α-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation, J. Struct. Biol. 137 (2002) 54–64.
[8] M.A. Walsh, T.R. Schneider, L.C. Sieker, Z. Dauter, V.S. Lamzin, K.S. Wilson, Refinement of triclinic hen egg-white lysozyme at atomic resolution, Acta Cryst. D: Biol. Crystallogr. 54 (1998) 522–546.
[9] A. Wlodawer, L.A. Svensson, L. Sjölin, G.L. Gilliland, Structure of phosphate-free ribonuclease A refined at 1.26 Å, Biochemistry 27 (1998) 2705–2718.
[10] A. Moulin, J.H. Bell, R.F. Pratt, D. Ringe, Inhibition of chymotrypsin by a complex of ortho-vanadate and benzohydroxamic acid: structure of the inert complex and its mechanistic interpretation, Biochemistry 46 (2007) 5982–5990.
[11] D.C. Rees, M. Lewis, W.N. Lipscomb, Refined crystal structure of carboxypeptidase A at 1.54 Å resolution, J. Mol. Biol. 168 (1983) 367–387.
[12] D.E. Tronrud, A.F. Monzingo, B.M. Matthews, Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin, Eur. J. Biochem. 157 (1986) 261–268.
[13] S.E. Phillips, B.P. Schoenborn, Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin, Nature 292 (1981) 81–82.
[14] I.G. Kamphuis, K.H. Kalk, M.B. Swarte, J. Drenth, Structure of papain refined at 1.65 Å resolution, J. Mol. Biol. 179 (1984) 233–256.
[15] M.M. Teeter, Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 6014–6018.
[16] T.L. Blundell, J.E. Pitts, I.J. Tickle, S.P. Wood, C.W. Wu, X-ray analysis (1.4-Å resolution) of avian pancreatic polypeptide: small globular protein hormone, Proc. Natl. Acad. Sci. U.S.A. 78 (1981) 4175–4179.
[17] T. Takano, Refinement of Myoglobin and Cytochrome C, in: S.R. Hall, T. Ashida (Eds.), Methods and Applications in Crystallographic Computing., Oxford University Press, Oxford, 1984, pp. 262–272.
[18] W. Bode, O. Epp, R. Huber, M. Laskowsk, W. Ardelt, The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3), Eur. J. Biochem. 147 (1985) 387–395.
[19] T.C. Terwilliger, D. Eisenberg, The structure of melittin. I. Structure determination and partial refinement, J. Biol. Chem. 257 (1982) 6010–6015.
[20] W. Steigemann, E. Weber, Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution, J. Mol. Biol. 127 (1979) 309–338.
[21] B.W. Dijkstra, K.H. Kalk, W.G. Hol, J. Drenth, Structure of bovine pancreatic phospholipase A2 at 1.7 Å resolution, J. Mol. Biol. 147 (1981) 97–123.
[22] J.T. Bolin, D.J. Filman, D.A. Matthews, R.C. Hamlin, J. Kraut, Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolatereductase refined at 1.7 Å resolution. I. General features and binding of methotrexate, J. Biol. Chem. 257 (1982) 13650–13662.
[23] P. Nordlund, H. Eklund, Structure and function of the *Escherichia coli* ribonucleotidereductase protein R2, J. Mol. Biol. 232 (1993) 123–164.
[24] B. Lovejoy, D. Cascio, D. Elsenberg, Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF), J. Mol. Biol. 234 (1993) 640–653.
[25] A. Shaw, D.E. McRee, V.D. Vacquier, C.D. Stout, The crystal structure of lysin, a fertilization protein, Science 262 (1993) 1864–1867.
[26] A. Okamoto, T. Higuchi, K. Hirotsu, S. Karamitsu, H. Kagamiyama, X-ray crystallographic study of pyridoxal 5′-phosphate-type aspartate aminotransferases from *Escherichia coli* in open and closed form, J. Biochem. 116 (1994) 95–107.
[27] E.K. O'Shae, J.D. Klemm, P.S. Kim, T. Alber, X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, Science 254 (1991) 539–544.
[28] W.S. Horne, J.L. Price, J.L. Keck, S.H. Gellman, Helix bundle quaternary structure from α/β-peptide foldamers, J. Am. Chem. Soc. 129 (2007) 4178–4180.
[29] D.L. Akey, V.N. Malashkevich, P.S. Kim, Buried polar residues in coiled–coil interfaces, Biochemistry 40 (2001) 6352–6360.
[30] P.B. Harbury, T. Zhang, P.S. Kim, T. Alber, A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants, Science 262 (1993) 1401–1407.
[31] W.S Horne, J.L. Price, S.H. Gellman, Interplay among side chain sequence, backbone composition, and residue rigidification in polypeptide folding and assembly, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 9151–9156.
[32] M.K. Yadav, L.J. Leman, D.J. Price, C.L. Brooks, C.D. Stout, M.R. Ghadiri, Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution, Biochemistry 45 (2006) 4463–4473.
[33] M.W. Giuliano, W.S. Horne, S.H. Gellman, An α/β-peptide helix bundle with a pure β^3-amino acid core and a distinctive quaternary structure, J. Am. Chem. Soc. 131 (2009) 9860–9861.
[34] W.S. Horne, L.M. Johnson, T.J. Ketas, P.J. Klasse, M. Lu, J.P. Moore, S.H. Gellman, Structural and biological mimicry of protein surface recognition by α/β-peptide foldamers, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 14751–14756.
[35] S. Kumar, M. Bansal, Geometrical and sequence characteristics of α-helices in globular proteins, Biophys. J. 75 (1998) 1935–1944.
[36] Data obtained on 1st August 2011 from the RCSB Protein Data Bank.
[37] K.L. Peña, S.E. Castel, C. de Araujo, G.S. Espie, M.S. Kimber, Structural basis of the oxidative activation of the carboxysomalγ-carbonic anhydrase, CcmM, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 2455–2460.