J Mol Model (2013) 19:2901–2911
DOI 10.1007/s00894-013-1819-7

ORIGINAL PAPER

Description of local and global shape properties of protein helices

Zhanyong Guo · Elfi Kraka · Dieter Cremer

Received: 15 October 2012 / Accepted: 5 March 2013 / Published online: 27 March 2013

Abstract A new method, dubbed "HAXIS" is introduced to describe local and global shape properties of a protein helix via its axis. HAXIS is based on coarse-graining and spline-fitting of the helix backbone. At each Cα anchor point of the backbone, a Frenet frame is calculated, which directly provides the local vector presentation of the helix. After cubic spline-fitting of the axis line, its curvature and torsion are calculated. This makes a rapid comparison of different helix forms and the determination of helix similarity possible. Distortions of the helix caused by individual residues are projected onto the helix axis and presented either by the rise parameter per residue or by the local curvature of the axis. From a non-redundant set of 2,017 proteins, 15,068 helices were investigated in this way. Helix start and helix end as well as bending and kinking of the helix are accurately described. The global properties of the helix are assessed by a polynomial fit of the helix axis and the determination of its overall curving and twisting. Long helices are more regular shaped and linear whereas short helices are often strongly bent and twisted. The distribution of different helix forms as a function of helix length is analyzed.

Keywords Protein helices · Helix axis · Helix bending · Helix twisting · Helix kinking

Introduction

The helix is one of the most common and most regular secondary structure elements (SSEs) in proteins [1–3]. Its shape is determined by the right-handed (or left-handed) spiral conformation of the protein backbone, in which H-bonding between the peptide NH group and carbonyl group provides suitable stability of the structure. Helices can be described from a biochemical point of view (via the frequency of H-bonding links measured by the number of residues per helix turn), a mathematical one (via the dimensions of a circumscribed cylinder), or with the help of a geometrical approach utilizing the radius R of the helix turns, the pitch p of the helix, the rise a of the helix per residue, or the phase angle γ per residue, which are all related to the dihedral angles ϕ_i and ψ_i of the residues i (see Fig. 1). A global feature of an ideal helix is the helix axis, which is given by the axis of the circumscribed cylinder. This axis is associated with a vector pointing from the helix start to the helix end thus defining both the direction of the helix and its position in the three-dimensional (3D) structure of the protein.

The helix axis has been used in protein investigations for visualization purposes [4], in 3D-structure analyses [5], for the description of helix packing [6–10], or the comparison of protein structures [11, 12]. In most of these investigations a simplified linear model of the helix axis was assumed [13, 14] although helices and their axes can be bent, twisted, or even kinked [15]. There are a variety of mathematical procedures to define a linear helix axis (and the associated axis vector) using various fitting approaches [13, 14, 16–18].

More realistic representations of the helix axis have to consider a possible bending or other irregularities of the helix axis [6]. A linear model can only approximate the real form of an axis, and therefore can lead to a misinterpretation of helix form and helix stability. Helices in proteins can be flexible, thus giving a protein a larger possibility of adjusting to its environment and to external forces. Because of this, methods have been developed that determine the bending or curving of a helix as reflected by the shape of the helix axis. Among these methods, the local axis methods used in HBEND [15, 19], the programs HELANAL [20–23] and P-CURVE [24], the

Z. Guo · E. Kraka · D. Cremer (✉)
Department of Chemistry, Southern Methodist University,
3215 Daniel Ave, Dallas,
TX 75275-0314, USA
e-mail: dcremer@smu.edu

E. Kraka
e-mail: ekraka@gmail.com

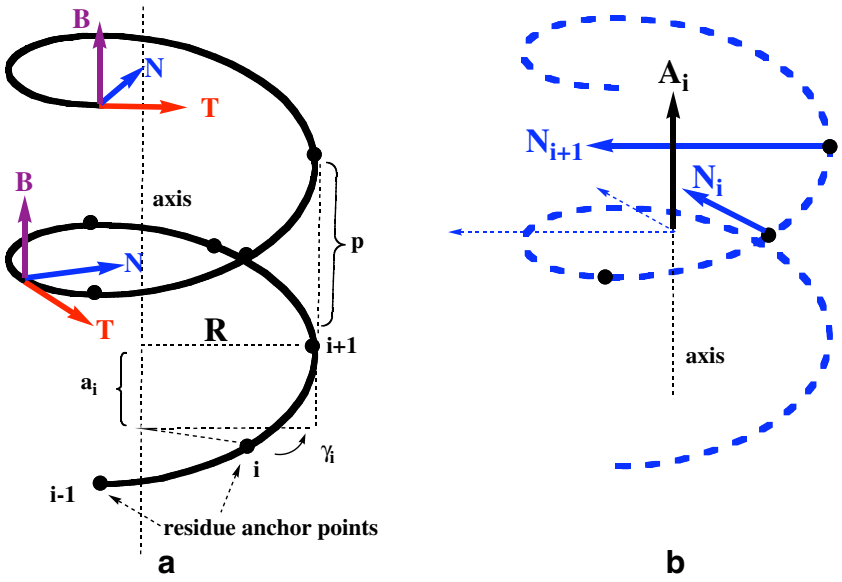

Fig. 1 **a** The helix of an ideal protein is represented as a smooth line determined by the positions of the *Cα* atoms (*dots*), which are the reference points of the spline-fitted backbone line. The following parameters of the helix are shown: *R* radius of the helix, *p* pitch, a_i rise parameter for residue *i*, γ_i corresponding phase angle. Two Frenet frames with unit tangent vector **T**, unit normal vector **N**, and unit binormal vector $\mathbf{B}_i$ are also shown. **b** The direction of the helix axis $\mathbf{A}_i$ at residue anchor point *i* is determined from the cross product of unit normal vectors $\mathbf{N}_i$ and $\mathbf{N}_{i+1}$ as indicated by the *dashed arrows*

QHELIX method [25] (based on the algorithms by Kahn [16, 26]), and the MC-HELAN algorithm [27] may be mentioned. Since the definition of the axis of a helix with just one or two turns is questionable, most of the algorithms developed consider helices with more than two turns. For example, HELANAL considers only helices with nine residues or more.

The current work is part of a project aimed at the computational description of proteins with methods based on coarse-graining, i.e., the atomistic presentation of the protein will be sacrificed for the benefit of obtaining the overall shape of a protein accurately. This approach requires different descriptions of a protein than those provided by Cartesian or internal coordinates. The overall shape of the latter can be described correctly, as we have shown in previous work [28], by using Frenet coordinates. In this work, we will apply this approach to helices to accomplish three objectives: (1) finding the right basis for a coarse-grained description of a protein helix; (2) determining the overall shape of a protein helix with accuracy without referring to an atomistic description; (3) rapidly comparing experimentally and computationally determined helix structures and evaluating helix similarity in a quantitative manner. The method presented here will be (among other similar approaches) the basis of parameterizing a suitable protein force field in a totally new way, which no longer depends on an atomistic approach.

Towards these objectives, we introduce a new definition and description of the helix axis, which combines four different features of a helix: (1) by a suitable projection technique we absorb all deviations from an ideal turn in the helix axis; (2) we quantify the local bending and wiggling of the helix axis by its curvature and torsion; (3) with the help of a polynomial fit, we smooth the helix axis obtained and then determine its overall bending and twisting; (4) finally, we determine the overall length and direction of the helix axis as it changes from the start to the end. This procedure, dubbed the HAXIS method, is applied to a set of 2,017 proteins with more than 15,000 helices. In this way, we obtain a detailed description of the (ir)regularities of helices, a clear definition of helix start and helix end, a detailed insight and a classification of helix bending, twisting, and even kinking, and a way of separating coils and short helices from normal helices. Since the axis is developed residue by residue to absorb irregularities in the residue conformations, and by this the turns of the helix, it is appropriate to speak of a helix axis only when three (or more) residues are present. In this way, a method is obtained that can be applied to the description of coils and very short helices as well as normal helices.

Computational procedures

In previous work, we have used Frenet frames and the associated scalar parameters curvature and torsion to describe the backbone of a protein [28]. For this purpose, we applied a coarse-grained representation of the backbone by choosing the *Cα* atoms of the residues as anchor points of the backbone, and converted the latter into a continuous smooth line in 3D-space by connecting the positions of the *Cα* anchor points with the help of a cubic spline fit. Curvature and torsion values at the anchor points of the continuous backbone line adopt characteristic values for a given SSE, which in this way can be identified and characterized easily [28].

In Fig. 1a, the backbone line of a helix is shown. A Frenet frame consisting of unit tangent vector **T**, unit normal vector **N**, and unit binormal vector **B** is depicted at the position of an anchor point of a residue. Vectors **T**, **N**, and **B** are

determined with the program APSA (Automated Protein Structure Analysis) [29] in the course of the calculation of curvature and torsion for each *Cα* of the protein backbone. Once the three Frenet vectors are available, calculation of the helix axis is straightforward and can proceed according to two different procedures.

In the first, the cross product of the unit normal vectors $\mathbf{N}_i$ and $\mathbf{N}_{i+1}$ of subsequent residues i and $i+1$ is taken thus yielding a vector $\mathbf{A}_i$ (Fig. 1b), which determines the direction of the helix axis at residue i. The subsequent cross product $\mathbf{N}_{i+1} \times \mathbf{N}_{i+2}$ yields the direction vector $\mathbf{A}_{i+1}$, which, for an ideal helix, coincides with the axis vector $\mathbf{A}_i$. For a real (non-regular) helix, the direction of the latter deviates slightly from that of the former, thus reflecting deviations in the structure of the helix as they are caused by conformational irregularities of individual residues. In this way, all distortions of the turns of a real helix caused by individual residues are reflected by the changes in the direction of the helix axis. Maximally, there can be $m-1$ changes for a helix consisting of m residues. Although it is of little use to speak of an axis in a global sense when considering the first and second turn of a helix, the helix axis as a local descriptor of irregularities in the residue-by-residue buildup of the helix is meaningful from the first residues on and therefore will be used in this work.

Despite the fact that the calculation of the helix axis via the cross products of the unit normal vectors is straightforward, it will lead to small errors in the helix direction if the angle between $\mathbf{N}_i$ and $\mathbf{N}_{i+1}$ becomes small. A mathematically more stable procedure of the piecewise calculation of the helix axis is obtained by utilizing the unit vectors **T** and **B** of a Frenet frame and obtaining the axis vector **A** from vector addition rather than a cross product. For the purpose of clarifying the basis of this second procedure, inspect Fig. 2a,b.

When looking from the top of a helix along its axis and shifting all unit binormal vectors $\mathbf{B}_i$ so that their starting points are identical to the origin of a sphere with unit radius, the end points of the vectors $\mathbf{B}_i$ of all residues form a circle (squares in Fig. 2a) on the surface of the sphere. The same is true if the starting points of the unit tangent vectors $\mathbf{T}_i$ are moved to the origin of the sphere (red points in Fig. 2a). In the case of an ideal helix, the center of the binormal circle and the tangent circle coincide and determine the direction of the helix axis (Fig. 2a). In case of a real helix with irregularities, the axis direction can change from residue i to residue $i+1$. This is considered by determining the changes in the directions of vectors $\mathbf{B}_i$ and $\mathbf{T}_i$ from residue i to residue $i+1$. In the case of an α-helix, the changes in the orientation angles will be close to 90° where $\mathbf{T}_i$ and $\mathbf{B}_i$ define one plane (shown in Fig. 2b) and vectors $\mathbf{B}_{i+1}$ and $\mathbf{T}_{i+1}$ a second plane (not shown in Fig. 2b), which is almost orthogonal to the first.

One can show that, for the distances between end points $\mathbf{B}_i$ and $\mathbf{B}_{i+1}$ ($\mathbf{T}_i$ and $\mathbf{T}_{i+1}$), i.e., $b_i = |\mathbf{B}_{i+1} - \mathbf{B}_i|$ and $t_i = |\mathbf{T}_{i+1} - \mathbf{T}_i|$ relationship (1) is fulfilled:

$$\frac{r_b}{r_t} = \frac{b_i}{t_i} \tag{1}$$

where r_b and r_t are the radii of the circles shown in Fig. 2b as defined by two residue points n and $n+1$. The local direction of the axis vector $\mathbf{A}_i$ can be calculated from Eq. (2):

$$\mathbf{A_i} = r_t\mathbf{B_i} + r_b\mathbf{T_i} \tag{2}$$

Using this procedure for all m residues of a helix, the total axis is the result of $m-1$ individual axis vectors, which in the case of a distorted helix point in slightly different directions and have slightly different lengths a_i. Distances a_i give always the shortest distance between two consecutive normals associated with the anchor points of the corresponding residues and, accordingly, provide a measure for the rise in the helix per residue i.

We have applied two different procedures to assess local distortions of the helix at position n. First, we used the calculated parameters a_i and compared them with the regular rise of the helix as found in an ideal helix of the α, 3_{10} or π-type (Table 1) [24]. Then, we utilized cubic spline fitting to convert the zigzagging axis line of a real (distorted) helix into a smooth line, for which we determine torsion and curvature at each residue anchor point. Both approaches provide sensitive measures for local distortions of the helix as caused by the nature of the residue in question.

To investigate the overall shape of a helix as caused by bending, torsion, or even kinking, we utilized a second order polynomial fit of the helix axis to obtain a smooth line with overall curving and well-defined direction vectors at its start and end points. In addition, third order polynomial fits were applied to test the non-planarity of a bent helix axis via its torsion, which can be considered as a measure for axis curvature in three dimensions. Fifth order polynomial fits were used to describe those helices, with distinct kinks that split the helix into two parts. The polynomial fits make it possible to provide, beside the local, also the global features of a helix via the properties of its axis. This investigation was carried out for helices with at least two turns, i.e. seven residues or more, because the global features of a helix are meaningful only in these cases. The analysis of local helix features, however, was carried out from the third residue on.

The calculation of the helix axis and the various analysis methods of the axis (global and local; second, third and fifth order polynomial fits and cubic spline fitting; calculation of Frenet frames leading to curvature and torsion; calculation of rise parameters a_i, helix radius R, and pitch p; statistical analysis, graphic presentations) are summarized in the program HAXIS, which is part of the APSA program [29] and

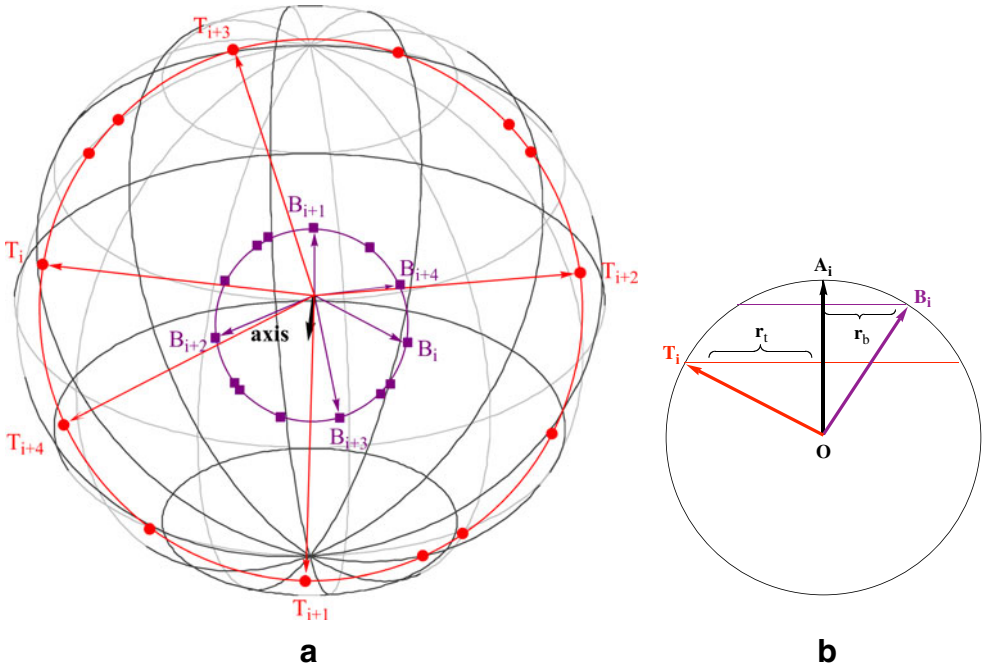

Fig. 2 **a** If the unit binormal vectors $\mathbf{B}_i$ of all anchor points of a helix are moved to a common origin, *O*, and the unit tangent vectors $\mathbf{T}_i$ to the same point, their end points lie on two circles on the surface of the circumscribed sphere with unit radius and origin *O*. **b** The associated vectors $\mathbf{T}_i$ and $\mathbf{B}_i$ lie on a slice through the sphere where the trace of the sphere is given as a unit circle and the diameters of the surface circles appear as secants of this unit circle. The parameters r_b and r_t give the radius of binormal and tangent circle, respectively, and make it possible to calculate the direction of the axis vector $\mathbf{A}_i$, which for an ideal helix is the same for all *i*. For a real helix, the surface circles are distorted and require the tangent-binormal pairs of residues i and $i+1$ (see text) for the calculation of a circle arc (i.e., the corresponding axis direction $\mathbf{A}_i$)

can be obtained via the APSA webpage from the authors (http://smu.edu/catco/apsa.html).

Results and discussion

A non-redundant dataset of 2,017 protein X-ray structures taken from the PDB [30] was selected, for which 15,068 helices with three or more residues were investigated. Of these, 11,761 helices possess seven residues or more. All protein structures considered had a resolution of 2 Å or better. Proteins with structure breaks or with ambiguous Cα positions were excluded from the analysis.

In Table 1, properties of ideal α, 3_{10} and π helices in polyalanines presented by smooth backbone lines are listed.

Table 1 Properties of ideal helices

Helix	Rise *a* [Å]	Radius *R* [Å]	Angle *γ* [°]	Residues per turn	Pitch *p* [Å]	*ϕ* [°]	*ψ* [°]
α	1.517	2.274	100.1	3.6	5.458	−57	−47[a]
3_{10}	1.955	1.868	121.5	3.0	5.793	−49	−26[a]
π	0.979	2.714	85.2	4.2	4.138	−57	−70[b]

[a] Ideal dihedral angles from Barlow and Thornton [15]

[b] Ideal dihedral angles from Armen and co-workers [38]

The backbone lines are fitted to the anchor point of each residue as given by the position of the *Cα* atoms. Rise parameter *a* per residue, radius *R*, pitch *p*, and phase angle *γ* per residue are clearly distinguishable for the three types of helices and clearly reflect the tighter winding of a 3_{10}-helix and the looser one of a π-helix (Table 1). For each of the ideal helices, the axis is a straight line with zero curvature and zero torsion, constant *a*, *p*, *R*, and *γ* values for each of the residues.

Global description of helix shape utilizing properties of its axis

In Table 2, the properties of some selected examples of real helices in proteins (reflected by their axes) are listed. They can be divided into three groups according to the bending of the helix axis. The bending is characterized by the average curvature κ_{av}, the maximum and minimum curvature $\kappa(\max)$ and $\kappa(\min)$ of the helix axis, the variation in the curvature $\Delta\kappa = \kappa(max) - \kappa(min)$, and the ratio $\eta_k = \Delta\kappa/\kappa_{av}$ (κ values always in units of $Å^{-1}$). Since the curvature values, apart from kinking situations, are rather small, it is easier to compare values $C(s) = 1/\kappa(s)$ where $C(s)$ is the radius of the osculating circle (associated with the curving of the helix axis) at an axis position defined by the arclength *s*. In this way parameters C_{av}, $C(\max)$, $C(\min)$, $\Delta C = C(max) - C(min)$,

Table 2 Properties of real helices calculated with HAXIS and compared to those obtained with HELANAL and HBEND. Not all digits are given for the curvature values κ_{av} to obtain the precise η_k and C values

Protein (PDB)[a]	Helix residues	κ_{av} [Å^{-1}]	$\Delta\kappa$ [Å^{-1}]	$\eta_k = \Delta\kappa/\kappa_{av}$	C_{av} [Å]	ΔC [Å]	$\eta_C = \Delta C/C_{av}$	Comp.[b] C [Å]	Type
1RIB	102–129	0.003	0.0001	0.05	333	17	0.05	140	Linear
1BGE	144–169	0.013	0.0020	0.15	78	12	0.16	73	Curved
1LIS	44–74	0.030	0.0140	0.47	34	19	0.56	25	Kinked
1LIS	44–61	0.026	0.0030	0.12	39	5	0.13		Curved
1LIS	62–74	0.023	0.0020	0.08	43	8	0.19		Curved
2TMN	281–295	0.008	0.0012	0.16	129	20	0.16	100	Linear
9PAP	25–42	0.006	0.0005	0.08	167	14	0.08	78[c]	Linear
1MBD	101–118	0.003	0.0001	0.03	311	11	0.03	184	Linear
1BP2	3–11	0.008	0.0002	0.03	131	4	0.03	53[c]	Linear
1BP2	39–54	0.008	0.0014	0.16	119	19	0.16	112	Linear
1BP2	89–106	0.006	0.0001	0.02	163	3	0.02	71[c]	Linear
5CYT	87–100	0.005	0.0006	0.12	198	22	0.12	94	Linear
5CPA	174–186	0.007	0.0007	0.10	156	15	0.10	71[c]	Linear
4LZT	25–35	0.025	0.0016	0.06	41	3	0.07	58	Curved
1MBD	124–149	0.011	0.0007	0.06	91	6	0.07	88	Curved
7RSA	24–33	0.048	0.0090	0.19	24	4	0.20	34	Curved
1MBD	59–77	0.013	0.0004	0.03	77	2	0.01	85	Curved
5CPA	15–27	0.014	0.0020	0.14	71	10	0.15	62	Curved
2OVO	34–43	0.020	0.0008	0.04	49	1.5	0.03	54	Curved
2TMN	136–151	0.033	0.0050	0.15	31	5	0.16	35[c]	Curved
5CPA	72–89	0.034	0.0110	0.32	31	11	0.37	34[c]	Kinked
1MBD	82–94	0.025	0.0170	0.66	41	27	0.66	24	Kinked

[a] 1RIB: ribonucleotide reductase protein R2 [39]; 1BGE: canine and bovine granulocyte-colony stimulating factor (G-CSF) [40]; 1LIS: Lysin [36]; 2TMN: thermolysin [41]; 9PAP: papain [42]; 1MBD: oxymyoglobin [43]; 1BP2: bovine pancreatic phospholipase [44]; 5CYT: Cytochrome C [45]; 5CPA: carboxypeptidase A [46]; 4LZT: egg-white lysozyme [47]; 7RSA: ribonuclease A [48]; 2OVO: silver pheasant ovomucoid [49]

[b] Comparative (Comp.) C values from Bansal and co-workers (HELANAL, first part of table) [21] and Barlow and Thorton (HBEND, second part of table) [15]

[c] Indicates a deviation from the classification obtained with HAXIS and given in the last column

and $\eta_C = \Delta C/C_{av}$ (C is always given in Å) are determined. The values listed in Table 2 are representative for 11,761 helices with seven or more residues. The distribution of the κ_{av} and $\Delta\kappa$ values of the 11,761 helices are shown in form of bar diagrams in Figs. 3a and 4a, respectively.

Utilizing calculated C_{av} and η_C values of the helix axis, protein helices with different forms of distortion can be classified in the following way.

(1) Linear and quasilinear helices have C_{av} values larger than 100 Å, for example 333 in the case of 1RIB (Table 2). About 18 % (2,155 helices) of all investigated helices fall into this class. According to Fig. 3, these are especially long helices with more than 13 residues. These helices are quite regular as is reflected by a low value of the ratio $\eta_C \leq 0.2$ and a small variation in rise per residue values a_i and pitch values P.

(2) Helices with a moderately bent axis possess C_{av} values between 30 and 100 Å ($0.01 \leq \kappa_{av} \leq 0.03$ Å^{-1}) represent the majority of all helices: about 54 % or 6,321 helices. In this group, there are especially helices with an average length of 12.7 residues (predominantly 12, up to as many as 33 residues).

(3) Helices with strongly bent or curved axes possess C_{av} values smaller than 30 Å ($\kappa_{av} \geq 0.03$ Å^{-1}) and represent the second largest group of all helices: about 28 % or 3,285 helices. In this group, there are short helices with an average length of 10.4 residues (predominantly 7 to 11 residues). Bending of the helix axis is mostly not regular and therefore one should speak of differently curved axes.

The regularity of a helix is reflected by the ratio η_K or η_C derived from the range of κ or C values and divided by the corresponding averages. These regularity descriptors do not necessarily correlate with the curving and twisting of the helix axis in the way that large curvature (torsion) also implies large irregularities. We find, however, a relationship between relatively long, (quasi)linear or moderately bent helices, which predominantly show high regularity as

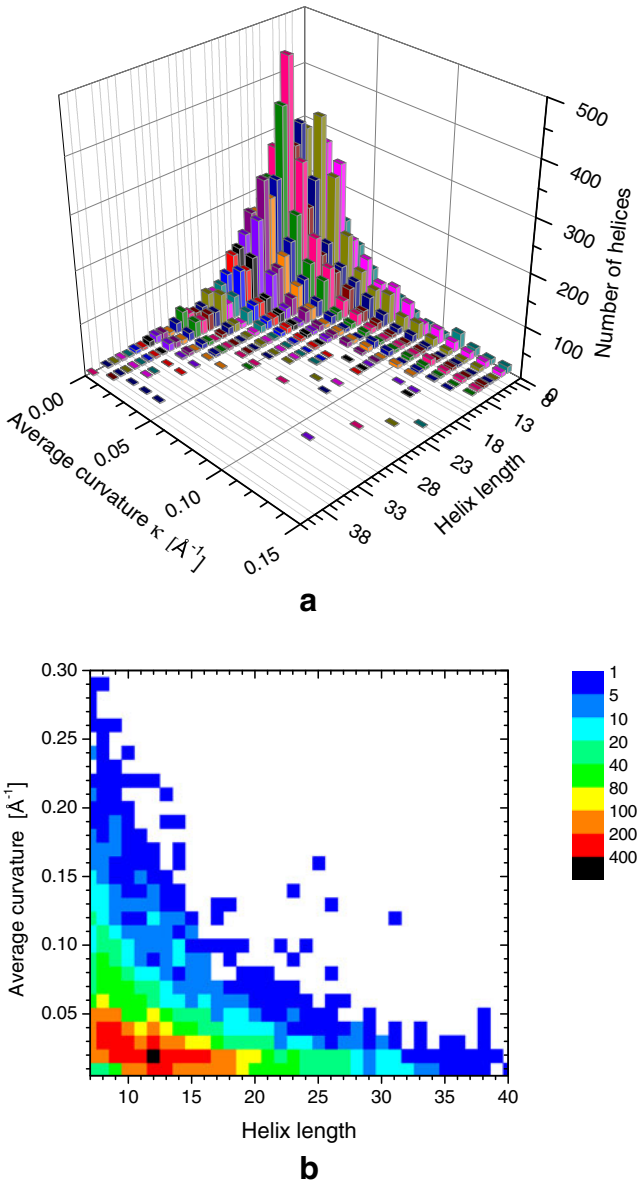

Fig. 3 Average curvature κ_{av} in $Å^{-1}$ of protein helices as a factor of helix length (i.e., number of residues) given (**a**) as general distribution in form of a 3D bar diagram (color is used to distinguish between helices of length m and $m+1$) and (**b**) in a detailed 2D distribution diagram where the frequencies of axis curvatures for specific helix lengths is color-coded. The curvature of the helix axis is determined after fitting all axis directions with a third-order polynomial. Only helices with seven or more residues are considered

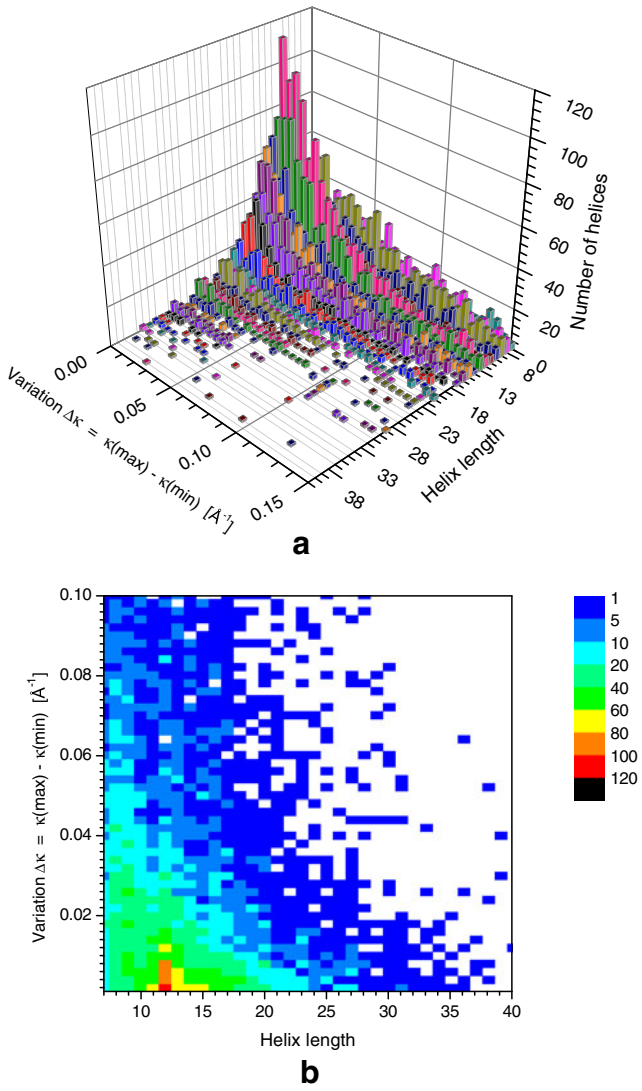

Fig. 4 Variation of the curvature $\Delta\kappa = \kappa(max) - \kappa(min)$ in $Å^{-1}$ of protein helices depending on helix length (i.e., number of residues) given (**a**) as general distribution in form of a 3D bar diagram (*color* is used to distinguish between helices of length m and $m+1$) and (**b**) in a detailed 2D distribution diagram where the frequencies of $\Delta\kappa$ values for specific helix lengths is color-coded. The curvature of the helix axis is determined after fitting all axis directions with a third-order polynomial. Only helices with seven or more residues are considered

reflected by $\eta_C \leq 0.15$. This group comprises 50 % (5,889) of all helices.

The group of helices with moderate irregularities ($0.15<\eta_C<0.35$) comprises only 24 % (2,857) of all helices and involves both moderately and strongly bent helices. Finally, there is a group of helices with low regularity ($\eta_C \geq 0.35$), the members of which are especially found among the short and strongly curved helices (26 % corresponding to 3,015 helices).

The curving of an axis does not always occur in a plane, as shown in Fig. 5 for a medium-sized helix. Hence, the torsion of the helix axis is larger than zero and can adopt values as large as 0.08 $Å^{-1}$ (corresponding to a torsion radius of 12.5 Å) as in the case shown in Fig. 5. To the best of our knowledge, the torsion of protein helices has not yet been investigated. The analysis of the average torsion τ_{av} of the helix axis reveals that the majority of bent helices is also twisted: more than 90 % of the helix axes possess a significant torsion value. About 20 % of them have low torsion values ($\tau_{av}<0.04$ $Å^{-1}$), 60 % have values $0.04<\tau_{av}<0.28$ $Å^{-1}$, and 20 % values $\tau_{av}>0.28$ $Å^{-1}$, i.e. high torsion. Only

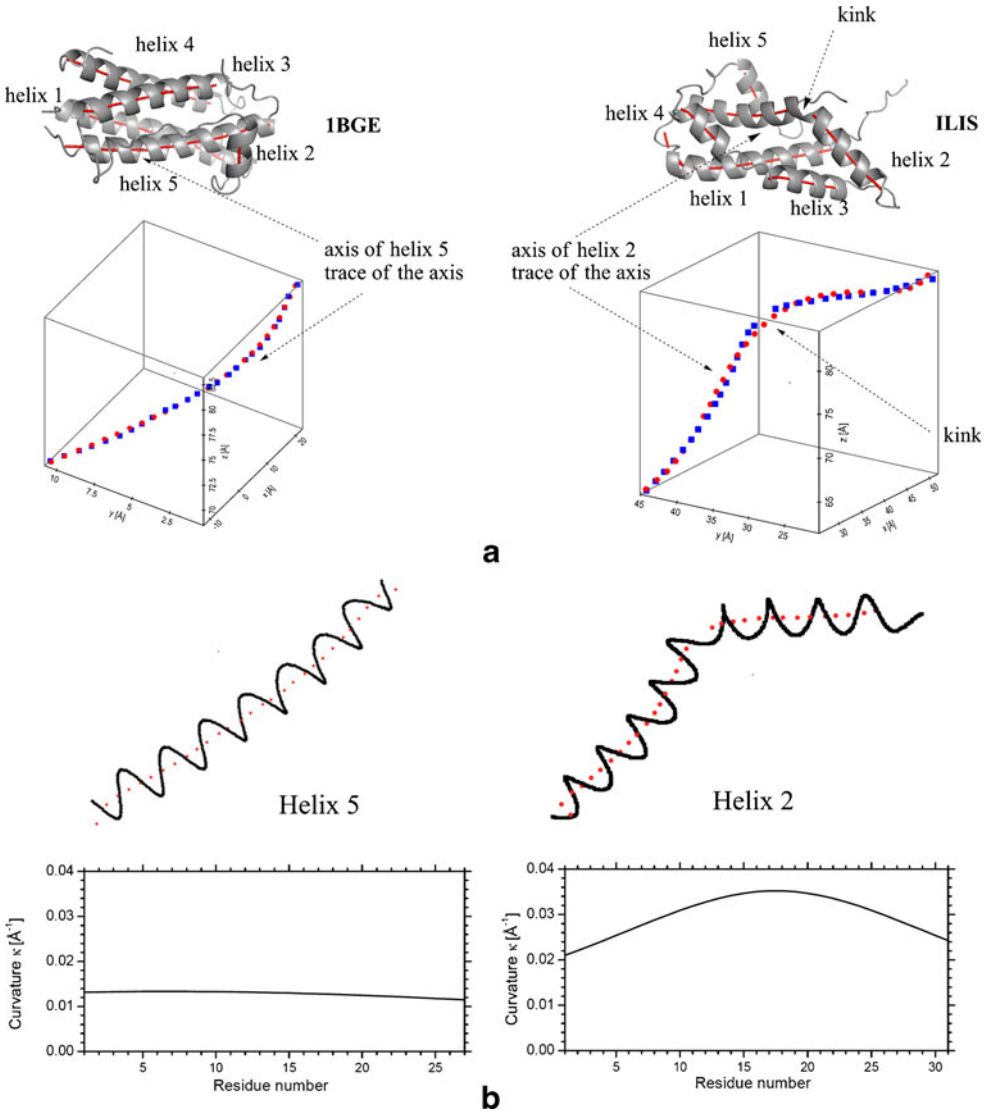

Fig. 5a,b Helix *n* of protein 1BGE {canine and bovine granulocyte-colony stimulating factor (G-CSF) [40] (*n*=5; left)} and protein 1LIS {Lysin [36]; (*n*=2; right).} **a** *Top*: Ribbon presentation. The calculated axes are given for all helices by *red tubes*. *Bottom*: Third order (1BGE, helix 5) and fifth order (1LIS, helix 2) polynomial fit (*red points*) for the calculated axis points (*blue points*). The perspective drawing provides an impression of the bending and twisting of the helix axis, which is not comprehended correctly when using a projection of a helix axis into a plane. **b** *Top*: Helix *n* and trace of its fitted axis (*red dots*). *Bottom*: Calculated curvature of the axis of helix *n* (based on a third or fifth order polynomial fit of the helix axis) as a function of the helix length (counting the first residue of helix *n* as residue 1)

9 % of all helices possess an axis, which bends in a plane ($\tau_{av} < 0.01$ Å^{-1}).

The extensive twisting of the helix axis indicates the limitations of any planar model of helix axis. It is interesting to note in this connection that the position of largest curving of the axis does not possess a significant torsion value and vice versa, i.e., in the majority of all cases first order curvature (curving in a plane) and second order curvature (twisting of the axis out of a plane and leading to the torsion) are related to each other in a helix.

Identification of kinked helices

Kinks are very common in transmembrane helices and have attracted much interest [31, 32]. The identification of a helix kink via H-bonding is often impossible because it requires a detailed investigation of helix geometry [27]. In this study, we identify helix kinking by both the global and the local description of helix shape via its axis. According to the global descriptors of the helix axis, helix kinking is given when C_{av} is about 40 or less (moderate to strong bending) and complemented by a irregularity descriptor η_C larger than 0.35. In this way, the kinked helices of Table 2 are identified. We note that helix kinking splits one helix into two helices and thereby affects the helix count. Also, a kinked helix has a different functionality in protein structure than a normal helix and in addition indicates a strong internal force [27, 33–35]. Therefore, the reliable identification of kinked helices is one of the objectives of protein structure analysis [27, 33].

Comparison with other descriptions of the helix axis

The HELANAL [20–22] and the HBEND method [15] are two established approaches to describe the shape of the helix axis. Common to both methods is that, after an appropriate calculation of the helix axis, the bending of the helix is

assumed to be regular and to take place in a plane so that a fit of the helix axis to a circle of radius R is sufficient. The current analysis reveals that such an assumption is an oversimplification and therefore it is not surprising that the shape description of helices obtained in this work differs from that given by HELANAL and HBEND. Some examples are listed in Table 2.

Deviations are found for all *C* values where in some cases (indicated by a c-superscript in Table 2) deviations are so large that they suggest a misleading classification of the shape of the helix axis. The helix examples of 9PAP, 1BP2 (residues 3–11), and 1BP2 (residues 89–106) are described by HBEND as curved because of predicted *C* values between 50 and 80 Å. According to HAXIS, the average *C* values (167, 131, 163, Table 2) are significantly above 100 Å, thus identifying these helices as typical examples of linear helices. Similarly, residues 136–151 in 2TMN form a helix that is described by HBEND and HELANAL as being kinked whereas C_{av} and η_C (31, 0.16, Table 2) are typical of a curved rather than kinked helix.

Local description of helix winding utilizing properties of its axis

In Fig. 6, the rise parameters a_i for all residues along a part of protein Lysin (PDB id: 1LIS) (from residue 5 to residue 133) [36] are shown according to residue number. As given in Table 1 for ideal helices, the *a*-parameter is always below 2 Å for helices and, accordingly, five helices can be identified (helix 1: residues 13–37; helix 2: 44–74; helix 3: 82–95; helix 4: 99–107; helix 5: 116–123; Fig. 6). These are interrupted by six coils, which have *a* values typically larger than 2. Hence, the *a*-parameters can be used effectively to determine the start and the end of a helix and to describe coils. However, they are less useful for the description of β-strands and turns, for which curvature and torsion of the backbone line [28] provide a better means. Hence, we have limited the use of the rise parameters *a* to helices and closely related SSEs.

Table 3 Identification of kink position using both maximum rise parameters *a*(max) and maximum curvature values of the spline-fitted helix axis

Protein (PDB) [a]	Residues	*a*(max) [Å]	Kink position[b]	Residues involved in kink[c]
1LIS	44–74	3.16	61	60–63
1MBD	82–94	2.58	86	85–88
1ECA	52–72	2.19	63	62–65
1A6M	83–95	2.66	86	85–88
1A8H	379–396	2.44	386, 391	385–388, 390–393
1AK0	236–263	2.94	241	240–243
1AH7	206–242	2.93	216	215–218
1ADE	183–199	3.22	193	192–195
1B8O	257–280	3.05	266	267–269
1C3D	20–37	2.43	25	24–27

[a] 1ECA: erythrocruorin [50]; 1A6M: myoglobin [51]; 1A8H: *Thermus thermophilus* methionyl-tRNA synthetase [52]; 1AKO: p1 nuclease [53]; 1AH7: phospholipase C [54]; 1ADE: adenylosuccinate synthetase [55]; 1B8O: purine nucleoside phosphorylase [56]; 1C3D: human C3d [57]

[b] According to *a*(max) values

[c] According to curvature values

In the center of helix 2 shown in Fig. 6, a single *a* value of 3.16 Å associated with residue 61 of 1LIS denotes the presence of a kink (see also Table 3) as correctly identified in the global description of the axis of helix 2 (Table 2, entry 1LIS; see also Fig. 5a, right side). The curvature diagram based on the spline-fitting of the axis points provides more details of this kink (Fig. 7), which effects residues 60–63. As shown in Table 3, the position of a kink can be identified easily for a helix utilizing the rise parameters a_i, which is unusually long between the two residues positioned directly at the kink position, thus reflecting the extra-strong bending of the helix axis (Fig. 5b, right). The curvature of any smooth line, which depends on the second derivative of the axis line with regard

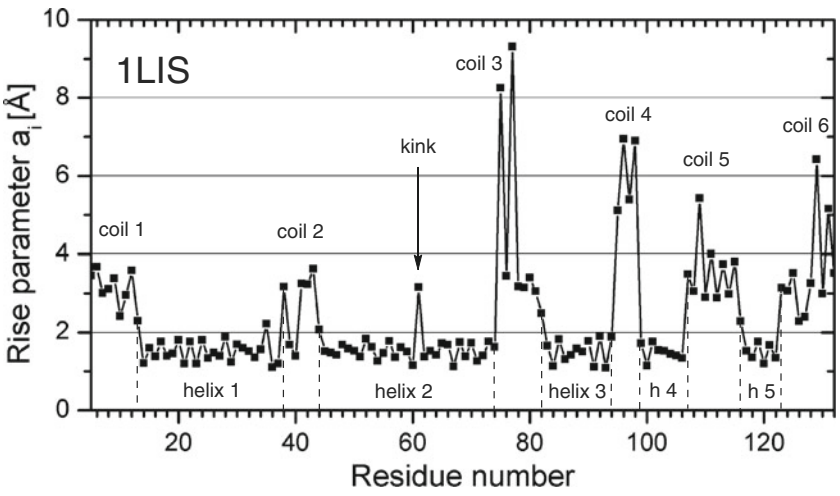

Fig. 6 Representation of the rise parameter a_i for protein 1LIS as a function of the residue number. Five helices (h) and six coils can be identified (*dashed lines* give start and end of helix). Residue 61 in helix 2 gives the position of a kink. For identification of residues in helix 2, cf. Fig. 7b

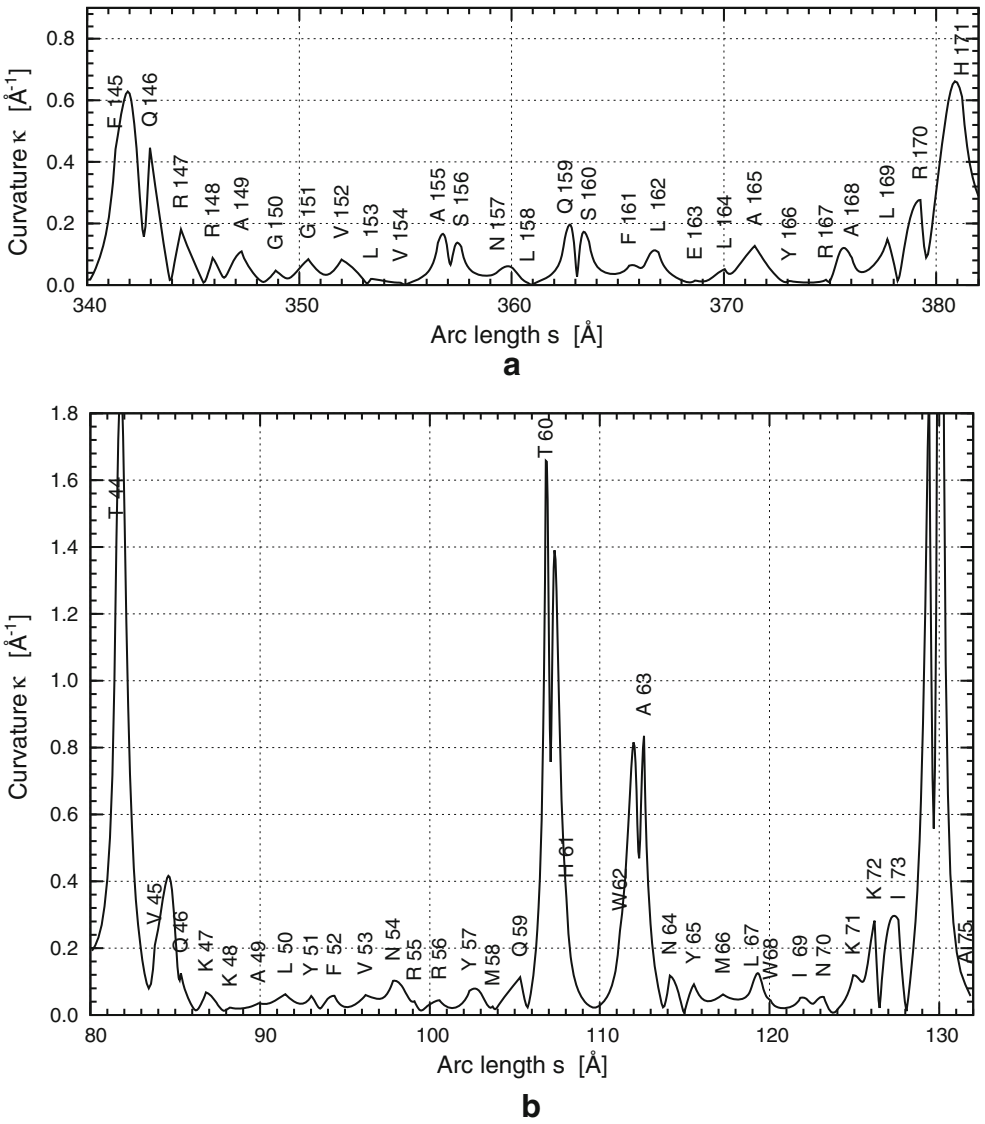

Fig. 7 Curvature diagram of the spline-fitted axis of helix *n* in (**a**) protein 1BGE (n=5) and (**b**) 1LIS (n=2). Bending as well as kinking involves always several residues: A155–S160 in 1BGE and T60–A63 in 1LIS. Note that at the ends of the helix the transition to a coil or turn causes an increase in curvature

to the arc parameter *s* [37] is more sensitive to kinking and, accordingly, leads to the quadruple-peak curvature pattern of Fig. 7b involving four residues (T60, H61, W62, A63 in 1LIS, Table 3). In this way, a kinking of the helix is identified easily as is shown in Table 3 for ten examples. HELANAL does not provide information about the exact kink position.

The curvature analysis reveals that curvatures peaks with $0.4 \leq \kappa \leq 1.0$ Å^{-1} indicate increasingly strong bending. Kinking is characterized by at least one curvature peak of the quadruple configuration being larger than 1.0 Å^{-1}. In this way, 708 kinked helices of a total of 11,761 helices investigated were identified, suggesting that, on average, 6 % of helices are kinked and an additional 4.8 % (564 helices) are strongly curved (0.6–1 Å^{-1}). Another 5.8 % (677 helices) shows small local irregularity with peak values of 0.4–0.6 Å^{-1}. Using the HELANAL method, Bansal and co-workers [21] predicted 4 % of all helices being kinked. The MC-HELAN method describes any non-linear helix as being kinked [27]. Obviously, one needs an accurate description of the helix axis (as used in this work) to detect all kinked helices and to distinguish them from strongly bent helices.

Most kinked helices have an average curvature value between 1 and 1.5 Å^{-1}. The number of even more strongly kinked helices decreases exponentially as shown in Fig. 8a. The majority of kinked helices are relatively short (8–16 residues, Fig. 8b). This observation is in line with the fact that long helices prefer a linear or quasilinear structure. Short helices are less stable and respond to forces exerted on the helix backbone first by bending and finally by kinking.

Comparison with other methods

Table 2 compares results obtained with HAXIS with those based on the programs HBEND [15, 19] and HELANAL

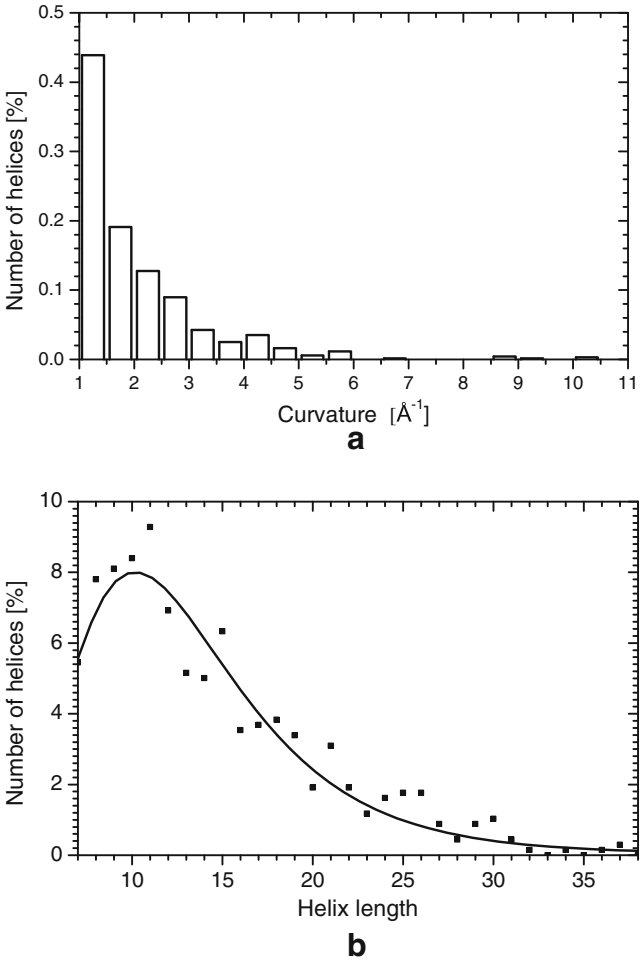

Fig. 8 Number of kinked helices in percent as a function of (**a**) average curvature values larger than 1 $Å^{-1}$ and (**b**) helix length

[20–23]. A comparison with other methods such as P-CURVE [24], QHELIX [25], or MC-HELAN [27] is difficult because these methods have objectives and/or use definitions that deviate from the helix description performed with HAXIS. A comparison with these methods is therefore beyond the scope of this investigation.

P-CURVE focuses on secondary structure assignments. In this connection, it can calculate the helix axis by a least-squares minimization based on all backbone atoms. In a complicated procedure, 14 parameters are determined that describe the helix. Only the circular bend of the helix axis can be calculated. The calculation of the axis curvature and torsion is not possible. QHELIX is based on a method suggested in the late 1980s by Kahn [16, 26] and determines inter-helical angles. QHELIX can also calculate the trace of the helix axis; however, it does not determine its curving or torsion. In addition, it requires as input start and end of the helix, which are automatically calculated by HAXIS. MC-HELAN focuses on the description of helix kinks where the definition of a kink deviates from that normally used in protein structure analysis (also used by HAXIS): for MC-HELAN, any non-linear, irregular part of the helix axis indicates a kink. The local helix axis is fitted by random sampling based on the distances of the backbone atoms to the helix axis. The axis is extended incrementally residue by residue. MC-HELAN describes helices as being either linear or kinked. The kinks are classified as bends (a change in the axis direction with all residues remaining helical) or disruptions (a change in the axis direction accompanied by a loss of helical character).

Conclusions

In this work, we have presented the HAXIS method, which provides a global and a local description of helix distortions via a global and a local definition of the helix axis. The axis is derived from a coarse-grained (a residue is represented by the $C\alpha$ atom as suitable anchor point) and spline-fitted backbone line of the helix provided by the APSA method [28]. At each anchor point of the helix the Frenet frame of the backbone line is calculated, which leads directly to a vector presentation of the helix axis: the conformation of each residue contributes to the winding of the helix and thereby to the axis direction. In the case of ideal helices, all local axis directions coincide and lead to one overall axis vector. In the case of a real helix, irregularities in the winding of the helix are reflected by different axis directions per residue and therefore a zigzagging axis line. For determination of the global shape of a helix, the axis line is smoothed by apolynomial fit of third order and its curving and twisting calculated.

The majority of all short helices (2–5 turns) is curved and twisted as is reflected by the distribution of average curvature and torsion. The shorter the helix, the more likely is a larger distortion from its regular shape. Utilizing the calculated curvature radius C, the variation ΔC, and the ratio $\eta_C = \Delta C/C_{av}$, helices are classified as being linear (18 %), moderately curved (54 %), or strongly curved (28 %). Long helices are preferentially linear and seem to be more resistant to distortions.

Both with the global and the local description, the kinking of a helix can be assessed quantitatively. Very reliable is the analysis of local curvature of the helix axis. At the position of a helix kink, a characteristic peak pattern (quadruple configuration of peaks) is found, which facilitates identification of kinking (curvature $\kappa > 1.0$ $Å^{-1}$) and strongly curved helices ($0.4<\kappa\leq 1.0$ $Å^{-1}$). On the average, 5.8 % of protein helices are kinked.

The characteristic curvature changes are used to identify the start and exit of a helix. Alternatively the rise parameters a_i can be used for this purpose or for the differentiation between helices and coils.

The approach presented in this work has several advantages compared to those previously published. The majority

of the latter methods is based on the assumption of a regular bending of the helix axis as described by a second-order polynomial ("circle fitting"), which leads to an oversimplification of the axis curvature radius *C*. Twisting of the helix axis can no longer be described in this way because bending as given by a second order polynomial takes place exclusively in one plane. HAXIS corrects the simplified helix presentations and quantifies axis torsion.

The HAXIS method described in this work is the basis for a rapid classification and comparison of calculated and measured helix structures and in this way can be used in connection with a coarse-grained description of proteins as will be described elsewhere.

Acknowledgments This work was supported financially by the National Science Foundation, Grant CHE 1152357. We thank Southern Methodist University for providing computational resources.

References

1. Richardson J (1981) Adv Protein Chem 34:167
2. Petsko G, Ringe D (2004) Protein structure and function. New Science, London
3. Pace C, Scholtz J (1998) Biophys J 75:422
4. Lopera J, Sturgis J (2005) J Mol Graph Model 23:305
5. Hu C, Koehl P (2010) Proteins 78:1736
6. Tatulian S (2008) Comput Biol Chem 32:370
7. Chothia C, Levitt M, Richardson D (1981) J Mol Biol 145:215
8. Walther D, Eisenhaber F, Argos P (1996) J Mol Biol 255:536
9. Lee S, Chirikjia G (2004) Biophys J 86:1105
10. Dalton J, Michalopoulos I, Westhead D (2003) Bioinformatics 19:1298
11. Singh A, Brutlag D (1997) Proc Int Conf Intell Syst Mol Biol 5:284
12. Gibrat J, Madej T, Bryant S (1996) Curr Opin Struct Biol 6:377
13. Åqvist J (1986) Comput Chem 10:97
14. Enkhbayar P, Damdinsuren S, Osaki M, Matsushima N (2008) Comput Biol Chem 32:307
15. Barlow D, Thornton J (1988) J Mol Biol 201:601
16. Kahn P (1989) Comput Chem 13:185
17. Christopher J, Swanson R, Baldwin T (1996) Comput Chem 20:339
18. Nievergelt Y (1997) Comput Aided Geomet Des 14:707
19. Blundell T, Barlow D, Borkakoti N, Thornton J (1983) Nature 306:281
20. Kumar S, Bansal M (1998) Biophys J 75:1935
21. Bansal M, Kumar S, Velavan R (2000) J Biomol Struct Dyn 17:811
22. Kumar S, Bansal M (1996) Biophys J 71:1574
23. Sugeta H, Miyazawa T (1967) Biopolymers 5:673
24. Sklenar H, Etchebest C, Lavery R (1989) Proteins 6:46
25. Lee H, Choi J, Yoon S (2007) Protein J 26:556
26. Kahn P (1989) Comput Chem 13:191
27. Langelaan D, Wieczorek M, Blouin C, Rainey J (2010) J Chem Inf Model 50:2213
28. Ranganathan S, Izotov D, Kraka E, Cremer D (2009) Proteins 76:418
29. Guo Z, Cremer D (2012) APSA12, Automated protein structure analysis. Southern Methodist University, Dallas, TX
30. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000) Nucleic Acids Res 28:235
31. Hall S, Roberts K, Vaidehi N (2009) J Mol Graph Model 27:944
32. Kauko A, Illergård K, Elofsson A (2008) J Mol Biol 380:170
33. Riek R, Rigoutsos I, Novotny J, Graham R (2001) J Mol Biol 306:349
34. Harris T, Graber A, Covarrubias M (2003) Am J Physiol Cell Physiol 285:C788
35. Reddy T, Ding J, Li X, Sykes B, Rainey J, Fliegel L (2008) J Biol Chem 283:22018
36. Shaw A, McRee D, Vacquier V, Stout C (1993) Science 262:1864
37. Kühnel W (2002) Differential Geometry: Curves-Surfaces-Manifolds. American Mathematical Society, Providence, RI
38. Armen R, Alonso D, Daggett V (2003) Protein Sci 12:1145
39. Nordlund P, Eklund H (1993) J Mol Biol 232:123
40. Lovejoy B, Cascio D, Eisenberg D (1993) J Mol Biol 234:640
41. Tronrud D, Monzingo A, Matthews B (1986) Eur J Biochem 157:261
42. Kamphuis I, Kalk K, Swarte M, Drenth J (1984) J Mol Biol 179:233
43. Phillips S, Schoenborn B (1981) Nature 292:81
44. Dijkstra B, Kalk K, Hol W, Drenth J (1981) J Mol Biol 147:97
45. Takano T (1984) Refinement of myoglobin and cytochrome *c*. In: Hall S, Hashida T (eds) Methods and applications in crystallographic computing. Oxford University Press, Oxford, pp 262–272
46. Rees D, Lewis M, Lipscomb W (1983) J Mol Biol 168:367
47. Walsh M, Schneider T, Sieker L, Dauter Z, Lamzin V, Wilson K (1998) Acta Crystallogr D: Biol Crystallogr 54:522
48. Wlodawer A, Svensson L, Sjolin L, Gilliland G (1988) Biochemistry 27:2705
49. Bode W, Epp O, Huber R, Laskowski M, Ardelt W (1985) Eur J Biochem 147:387
50. Steigemann W, Weber E (1979) J Mol Biol 127:309
51. Vojtechovsky J, Chu K, Berendzen J, Sweet R, Schlichting I (1999) Biophys J 77:2153
52. Sugiura I, Nureki O, Ugaji-Yoshikawa Y, Kuwabara S, Shimada A, Tateno M, Lorber B, Giege R, Moras D, Yokoyama S et al (2000) Structure 8:197
53. Romier C, Dominguez R, Lahm A, Dahl O, Suck D (1998) Proteins 32:414
54. Hough E, Hansen L, Birknes B, Jynge K, Hansen S, Hordvik A, Little C, Dodson E, Derewenda Z (1989) Nature 338:357
55. Silva M, Poland B, Hoffman C, Fromm H, Honzatko R (1995) J Mol Biol 254:431
56. Fedorov A, Shi W, Kicska G, Fedorov E, Tyler P, Furneaux R, Hanson J, Gainsford G, Larese J, Schramm V et al (2001) Biochemistry 40:853
57. Nagar B, Jones R, Diefenbach R, Isenman D, Rini J (1998) Science 280:1277