electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

Efficient characterization of protein secondary structure in terms of screw motions

Gerald R. Kneller and Paolo Calligari

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Efficient characterization of protein secondary structure in terms of screw motions

Gerald R. Kneller[a,b]* and Paolo Calligari[a,c]

[a]Laboratoire Léon Brillouin, CEA Saclay, 91191 Gif-sur-Yvette, France, [b]Centre de Biophysique Moléculaire, CNRS (affiliated with the University of Orléans and with INSERM), 45071 Orléans, France, and [c]Institut Laue–Langevin, 6 Rue Jules Horowitz, BP 156, 38042 Grenoble, France

Correspondence e-mail: kneller@cnrs-orleans.fr

Received 1 August 2005
Accepted 20 December 2005

A simple and efficient method is presented to describe the secondary structure of proteins in terms of orientational distances between consecutive peptide planes and local helix parameters. The method uses quaternion-based superposition fits of the protein peptide planes in conjunction with Chasles' theorem, which states that any rigid-body displacement can be described by a screw motion. The helix parameters are derived from the best superposition of consecutive peptide planes and the 'worst' fit is used to define the orientational distance. Applications are shown for standard secondary-structure motifs of peptide chains for several proteins belonging to different fold classes and for a description of structural changes in lysozyme under hydrostatic pressure. In the latter case, published reference data obtained by X-ray crystallography and by structural NMR measurements are used.

1. Introduction

The determination and characterization of protein secondary structure is a fundamental task in molecular biology, crystallography and simulation studies. In many situations the necessity arises to quantify particular structural changes of a protein that arise from a change in its environment. The influence of temperature or pressure on the fold of a protein is a typical example. Standard motifs in protein secondary structure are traditionally described in terms of two torsional angles, φ and ψ, per residue, which define for each C^{α} atom the rotation of the left and right peptide plane about the $N-C^{\alpha}$ and $C^{\alpha}-C$ bond, respectively (Stryer, 1988). In the past, various methods have been developed to determine secondary-structure elements (Kabsch & Sander, 1983; Richards & Kundrot, 1988; Frishman & Argos, 1995; Taylor, 2001) and to describe their geometry in more detail (Barlow & Thornton, 1988; Sklenar *et al.*, 1989; Thomas, 1994). A rigourous mathematical description of protein secondary structure can be obtained by applying the theory of screw motions, in which the winding of the protein backbone is described in terms of local helix parameters. The theory of screw motions goes back to the mathematician M. Chasles (Chasles, 1830, 1870) and a useful recent introduction can be found in the book by Selig (1996). In a recent paper, Quine uses screw-motion theory and constructs local helix parameters for a protein from the torsion angles φ and ψ (Quine, 1999). An important step is the introduction of quaternions, which can be related to the (φ, ψ) angles on one hand and to the rotation/helix axis on the other.

In this article, we present an efficient method for the characterization of protein secondary structure which is based on quaternion superposition fits of consecutive peptide planes. From the resulting quaternion parameters, we construct the

local helix geometry of the protein backbone and show that the superposition method may also be used to define a scalar measure for the orientational distance between consecutive peptide planes. The latter allows distinction between all common secondary-structure motifs, such as different helix types and β-strands, with the exception of handedness.

In the following section the method is briefly explained and applications are presented in §3. The first application concerns an illustration for simple model structures, such as right- and left-handed α-helices and β-strands. We show then how our method works for proteins which fall into different fold classes and finally discuss in more detail how it can be used to quantify changes in the secondary structure of lysozyme which are caused by external pressure. For this purpose, we use published reference structures which have been obtained from X-ray crystallography and from structural NMR measurements. The essential results are summarized and discussed in §4. In Appendix *A*, we recall the essential properties of quaternions and give a short constructive proof of Chasles' theorem, which demonstrates the usefulness of quaternion calculus.

2. Method

As stated in §1, our method for the description of protein secondary structure relies on quaternion-based superposition fits of molecular structures. The method is well established and we refer to articles by Kearsley (1989) and by Kneller (1991) for details. Here, we utilize the fact that the quaternion method not only yields the 'best' fit, from which local helix parameters describing the winding of the protein backbone can be constructed, but also the 'worst' fit, from which an orientational distance measure can be derived.

2.1. Quaternion superposition fits

Suppose that $\{\vec{r}_\alpha\}$ and $\{\vec{r}'_\alpha\}$ are two sets of vectors describing the positions of atoms representing equivalent molecular structures *A* and *B*, respectively. Both structures contain the same number of atoms and are somehow placed in space. A rigid-body displacement $A \rightarrow B$ can be defined as an optimization problem, where structure *A* is fitted onto structure *B* in a least-squares sense. In the case that both structures are identical, the resulting fit error will be zero. One starts by constructing the translation vector $\vec{t} = \vec{R}_{C'} - \vec{R}_C$ connecting the two centres of rotation, *C* and *C*′, which are to be chosen in the same way for *A* and *B*, and computes the coordinate sets $\{\mathbf{x}_\alpha\}$ and $\{\mathbf{x}'_\alpha\}$ containing the relative atomic positions to the respective rotation centres. Here and in the following the prime refers to the target structure *B*. The optimal rotation is obtained by minimizing the target function

$$m(q) = \sum_{\alpha=1}^{N} w_\alpha (\mathbf{D}\cdot\mathbf{x}_\alpha - \mathbf{x}'_\alpha)^2 \qquad (1)$$

with respect to a set of angular variables which parametrize the orthogonal rotation matrix **D**. Each atom is assigned a positive weight w_α, with $\sum_\alpha w_\alpha = 1$. A convenient set of angular variables are normalized (real) quaternion parameters, $q = \{q_0, q_1, q_2, q_3\}$, with $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. In this case **D** takes the form (Altmann, 1986)

$$\mathbf{D}(q) = \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(-q_0q_3+q_1q_2) & 2(q_0q_2+q_1q_3) \\ 2(q_0q_3+q_1q_2) & q_0^2+q_2^2-q_1^2-q_3^2 & 2(-q_0q_1+q_2q_3) \\ 2(-q_0q_2+q_1q_3) & 2(q_0q_1+q_2q_3) & q_0^2+q_3^2-q_1^2-q_2^2 \end{bmatrix} \qquad (2)$$

and describes a proper rotation with $\det(\mathbf{D}) = +1$. Using the orthogonality of **D**, the target function $m(q)$ can be written as a quadratic form in the quaternion parameters,

$$m(q) = \mathbf{q}^T \cdot \mathbf{M} \cdot \mathbf{q}, \qquad (3)$$

where $\mathbf{q} = (q_0, q_1, q_2, q_3)^T$ is a column vector and **M** is a positive semi-definite matrix. The superscript *T* denotes a transposition. The matrix **M** has the form (Kearsley, 1989; Kneller, 1991),

$$\mathbf{M} = \sum_{\alpha=1}^{N} w_\alpha \begin{bmatrix} (\mathbf{x}_\alpha - \mathbf{x}'_\alpha)^2 & \mathbf{u}_\alpha^T \\ \mathbf{u}_\alpha & \mathbf{P}_\alpha \end{bmatrix} \qquad (4),$$

where $\mathbf{u}_\alpha$ and $\mathbf{P}_\alpha$ are given by

$$\mathbf{u}_\alpha = \mathbf{x}_\alpha \wedge \mathbf{x}'_\alpha, \qquad (5)$$

$$\mathbf{P}_\alpha = \mathbf{x}_\alpha \cdot \mathbf{x}'^T_\alpha + \mathbf{x}'_\alpha \cdot \mathbf{x}^T_\alpha. \qquad (6)$$

The minimization of $m(q)$ with respect to the quaternion parameters must be performed with the side constraint $\mathbf{q}^T\cdot\mathbf{q} = 1$. Using the method of Lagrange multipliers, one is led to the eigenvector problem

$$\mathbf{M}\cdot\mathbf{q} = \lambda\mathbf{q}. \qquad (7)$$

Since $m(q) \geq 0$, the matrix **M** is positive semi-definite and one obtains a set of four real eigenvalues, $\{\lambda_j\}$, with $\lambda_j \geq 0$ ($j = 1, \ldots, 4$), and a set of corresponding orthonormal eigenvectors, $\mathbf{q}_j$, with $\mathbf{q}_j^T\cdot\mathbf{q}_k = \delta_{jk}$. Here, δ_{jk} is the Kronecker symbol. It follows from (3) and (7) that

$$m(q_j) = \lambda_j. \qquad (8)$$

The eigenvalues are thus the residuals of the fit and can be ordered such that

$$\lambda_1 \leq \lambda_2 < \lambda_3 \leq \lambda_4. \qquad (9)$$

The quaternion corresponding to the smallest eigenvalue, λ_1, is thus the solution for the optimal fit and the quaternion parameters q_1 describe the relative orientation of $\{\mathbf{x}'_\alpha\}$ with respect to $\{\mathbf{x}_\alpha\}$.

We note that one obtains two twofold-degenerate eigenvalues if the structures to be superposed are linear. In this case one has (Kneller, 1991)

$$\lambda_{a,b} = \sum_\alpha w_\alpha (|\mathbf{x}_\alpha|^2 + |\mathbf{x}'_\alpha|^2 \mp 2|\mathbf{x}_\alpha||\mathbf{x}'_\alpha|), \qquad (10)$$

where $a = 1, 2$, $b = 3, 4$ and both the rotations leading to the minimum and maximum distance are not uniquely determined. Any normalized linear combination of the two eigen-

vectors associated with λ_a and λ_b, respectively, describes an equivalent rotation.

The use of quaternion parameters in not only very convenient for finding a rigid-body transformation between two sets of coordinates, but the result can also be directly related to conventional representations of rotations. Here, the following relation is of importance:

$$\mathbf{q} \equiv \begin{pmatrix} q_0 \\ \mathbf{q}_v \end{pmatrix} = \begin{bmatrix} \cos(\varphi/2) \\ \sin(\varphi/2)\mathbf{n} \end{bmatrix}. \tag{11}$$

From the scalar part of a quaternion, q_0, one thus directly obtains the rotation angle; the rotation axis can be extracted from the vectorial part, $\mathbf{q}_v$. It should be noted that the transformation $\varphi \rightarrow \varphi + 2\pi$, which leaves the rotation matrix $\mathbf{D}(\mathbf{n}, \varphi)$ invariant, leads to a global change in the sign of the quaternion parameters. It can easily be verified that $\mathbf{q}(\mathbf{n}, \varphi + 2\pi) = -\mathbf{q}$. The pair of quaternions $\{Q, -Q\}$ is thus mapped onto the same rotation matrix $\mathbf{D}(q)$.

2.2. Orientational distance

The eigenvalue describing the 'worst' superposition (λ_4 according to the ordering scheme; equation 9) can be used to define an orientational distance between two molecular structures *via*

$$\Delta_\Omega = \left(\frac{M_{11}}{\lambda_4}\right)^{1/2}. \tag{12}$$

Equation (4) shows that the matrix element M_{11} contains the squared Euclidean distance between the vector sets $\{\vec{x}_\alpha\}$ and $\{\vec{x}'_\alpha\}$ and therefore Δ_Ω is the Euclidean distance normalized to its maximum possible value. Consequently,

$$0 \leq \Delta_\Omega \leq 1. \tag{13}$$

It is important to note that (12) yields a unique orientational distance of two linear molecular structures whose relative orientation has no unique description in terms of angular variables. Supposing that $|\mathbf{x}'_\alpha| = |\mathbf{x}_\alpha|$ for $\alpha = 1, \ldots, N$, we see from (10) that for linear rigid bodies $\Delta_\Omega = 0$ in the parallel configuration and $\Delta_\Omega = 1$ in the antiparallel configuration. We note here that λ_b, as given by (10), is a strict upper limit for the Euclidean distance of two molecular structures in general (Kneller, 2005).

2.3. Chasles' theorem

Let $\mathbf{r} = (x, y, z)^T$ be a column vector containing the coordinates of a radius vector $\vec{r} = \overrightarrow{OP}$ of a point P in a rigid body, where O is the origin of the coordinate system. An arbitrary rigid-body displacement is described by a rotation about a point C, which is not necessarily located inside the rigid body, and a subsequent translation. Let $\mathbf{R}_C$ be the coordinates of the radius vector $\vec{R}_C = \overrightarrow{OC}$ and let $\mathbf{t}$ be the coordinates of the translation vector $\vec{t} = \overrightarrow{CC'}$, where C' is the centre of rotation after the translation. The coordinates of P after a rigid-body displacement are then given by

$$\mathbf{r}' = \mathbf{R}_C + \mathbf{D} \cdot (\mathbf{r} - \mathbf{R}_C) + \mathbf{t}, \tag{14}$$

where $\mathbf{D}$ is an orthogonal 3×3 matrix. In the following, only proper rotations with $\det(\mathbf{D}) = +1$ will be considered. If $\mathbf{n} = (n_x, n_y, n_z)^T$ contains the components of the unit vector $\vec{n}$, pointing in the direction of the rotation axis, and φ is the angle of rotation, the corresponding rotation matrix can be written as

$$\mathbf{D}(\mathbf{n}, \varphi) = \mathbf{P}_\parallel + \cos(\varphi)\mathbf{P}_\perp + \sin(\varphi)\mathbf{N}, \tag{15}$$

where $\mathbf{P}_\parallel = \mathbf{n}\cdot\mathbf{n}^T$ and $\mathbf{P}_\perp = \mathbf{1} - \mathbf{P}_\parallel$ are the projectors onto $\vec{n}$ and its complement, respectively, and $\mathbf{N}$ is the antisymmetric matrix

$$\mathbf{N} = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix}. \tag{16}$$

The theorem of Chasles states that one can find a reference point X, whose radius vector $\vec{R}_x = \overrightarrow{OX}$ has the coordinates $\mathbf{R}_x$, such that

$$\mathbf{r}' = \mathbf{R}_x + \mathbf{D}(\mathbf{n}, \varphi) \cdot (\mathbf{r} - \mathbf{R}_x) + \alpha\mathbf{n}. \tag{17}$$

This coordinate transformation describes a screw motion with translation α parallel to the axis of rotation. For the following considerations, we introduce the difference vector $\vec{u} = \vec{R}_x - \vec{R}_c$. Equating (14) and (17) and using the fact that $\mathbf{n}$ is an eigenvector of $\mathbf{D}$, one finds that the coordinates of $\vec{u}$ satisfy the following set of linear equations

$$(\mathbf{1} - \mathbf{D}) \cdot \mathbf{u} = \mathbf{t}_\perp. \tag{18}$$

Here, $\mathbf{t}_\perp = \mathbf{P}_\perp\cdot\mathbf{t}$. As shown in Appendix B, the above equation has a linear manifold of solutions,

$$\mathbf{u}(\lambda) = \mathbf{u}_\perp + \lambda\mathbf{n}, \quad \lambda \in \mathbb{R}, \tag{19}$$

where $\mathbf{u}_\perp$ is perpendicular to $\mathbf{n}$ and has the explicit form

$$\mathbf{u}_\perp = \tfrac{1}{2}[\mathbf{t}_\perp + \cot(\varphi/2)\mathbf{n} \wedge \mathbf{t}]. \tag{20}$$

In absolute coordinates, the axis of the screw motion is given by

$$\mathbf{R}_x = \mathbf{R}_C + \mathbf{u}_\perp + \lambda\mathbf{n}, \tag{21}$$

and

$$\mathbf{R}_x^\perp = \mathbf{R}_C + \mathbf{u}_\perp \tag{22}$$

contains the coordinates of the radius vector $\vec{R}_x^\perp$ relating the origin to the point $X^\perp$ on the helix axis which is closest to the reference point C. In the following, $X^\perp$ will be referred to as the centre of screw motion. The radius ρ of the corresponding screw motion is given by the Euclidean length of $\vec{u}_\perp$, since the latter is the vector pointing from $\vec{R}_C$ to $X^\perp$. Using (20) one finds

$$\rho = \frac{|\mathbf{t}_\perp|}{2}[1 + \cot^2(\varphi/2)]^{1/2}. \tag{23}$$

It should be noted that ρ diverges if φ is a multiple of 2π, corresponding to pure translations, and if $|\mathbf{t}_\perp| \neq 0$.

3. Applications

3.1. Screw-motion description of protein main chains

The method described above, which will be referred to as *ScrewFit* in the following, is now applied to define the local helical structure of polypeptides and proteins. The rigid bodies are here the triangles formed by the atoms (O, C, N) in the backbone of polypeptides (see Fig. 1) which define the so-called peptide planes. Here, the C atoms are the centres of rotation and the translation vectors are thus the position differences between the C atoms in consecutive amino acids, $\mathbf{t}_i = \mathbf{R}_{C(i+1)} - \mathbf{R}_{C(i)}$. The quaternion parameters $\mathbf{q}_i$ are obtained from the fit of the (O, C, N) triangle of peptide bond i onto that of peptide bond $i + 1$. From each set of quaternion parameters the direction $\vec{n}$ of the rotation axis and the rotation angle φ can be computed from (11).

The following parameters are used to define the local helix structure of a polypeptide.

(i) The helix radius ρ defined in (23).

(ii) The number of amino acids per turn,

$$\tau = (2\pi/\varphi). \tag{24}$$

(iii) The pitch, which is defined as

$$p = |\mathbf{R}^{\perp}_{x,i+1} - \mathbf{R}^{\perp}_{x,i}|\tau. \tag{25}$$

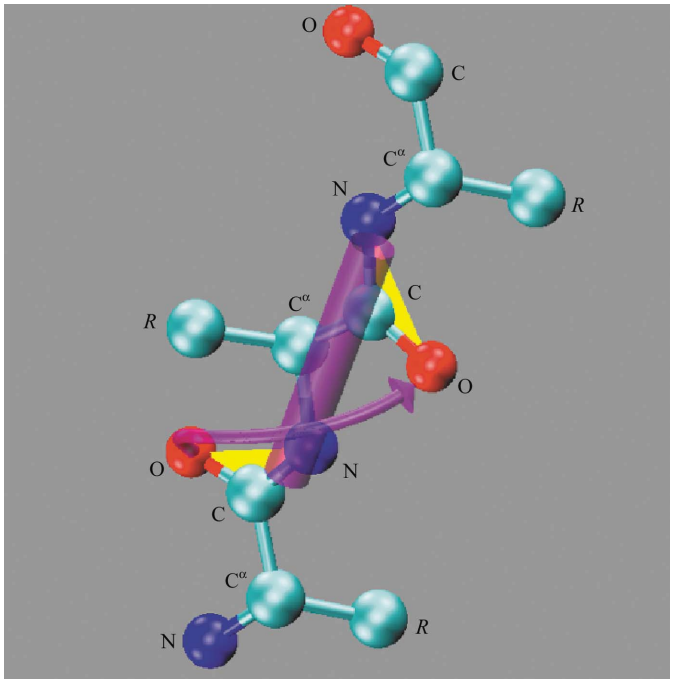

Figure 1
A tripeptide with two peptide bonds in the extended conformation, where *R* represents unspecified side chains. The screw motion relating the yellow triangles formed by the O, C, N atoms of the peptide planes defines the local helix, which is schematically represented by the cylinder in purple and the corresponding screw arrow. The radius of the cylinder corresponds to the radius of the screw motion.

Table 1
Helix parameters for different model structures.

Here, ρ is the helix radius with the C atom of the peptide plane on the helix surface, ρ_{C^α} is the corresponding radius if the C atom is replaced by the C^α atom, τ is the number of residues per turn, **h** is the handedness and σ the straightness parameter. The latter equals 1 for all model structures, since none is curved. The straightness parameter for the extended conformation cannot be defined, since the model structure for the latter consists of only three residues (see Fig. 1). Further explanations are given in the text.

Motif	ρ (nm)	ρ_{C^α} (nm)	τ	Pitch (nm)	h	σ	Δ_Ω
α-Helix (R)	0.171	0.227	3.62	0.556	+	1	0.582
α-Helix (L)	0.171	0.227	3.62	0.556	−	1	0.582
3_{10}-Helix	0.146	0.203	3.28	0.589	+	1	0.670
π-Helix	0.178	0.258	4.16	0.558	+	1	0.471
β-Strand	0.055	0.093	2.03	0.671	−	1	0.875
Extended	0.037	0.055	2.00	0.725	−		0.754

Here, $\vec{R}^{\perp}_{x,i}$ is the radius vector pointing from the origin to the centre $X^{\perp}_i$ of the screw motion relating peptide plane i and peptide plane $i + 1$.

(iv) The handedness, which is defined as the sign of the projection of the translation vector $\vec{t}_i$ onto the direction $\vec{n}_i$ of the local helix axis,

$$h = \mathrm{sign}(\mathbf{n}_i^T \cdot \mathbf{t}_i). \tag{26}$$

(v) The straightness parameter σ of the local helix axis. For residue i the latter is defined as

$$\sigma_i = \boldsymbol{\mu}_i^T \cdot \boldsymbol{\mu}_{i+1}, \tag{27}$$

where

$$\boldsymbol{\mu}_i = \frac{\mathbf{R}^{\perp}_{x,i+1} - \mathbf{R}^{\perp}_{x,i}}{|\mathbf{R}^{\perp}_{x,i+1} - \mathbf{R}^{\perp}_{x,i}|}. \tag{28}$$

(vi) The orientational distance between the peptide planes (O, C, N) in residues i and $i + 1$, which is defined through (12).

3.2. Model structures

We first apply *ScrewFit* to well known model structures of polypeptides which have been taken from the Image Library of Biological Macromolecules in Jena (Institute of Molecular Biotechnology, Jena; http://www.imb-jena.de/IMAGE.html). Table 1 shows the corresponding local helix parameters which have been defined in the previous section. All model peptides are polyalanine molecules containing ten residues, except for the extended conformation, which is represented by the alanine tripeptide shown in Fig. 1. In all cases the N-terminus is the starting point of the respective polypeptide chain. In the context of our study all motifs are considered as helices, a β-strand being simply a thin left-handed helix with two amino acids per turn.

The parameters concerning the different secondary structure motifs shown in Table 1 may be compared with those published in the study of Barlow & Thornton (1988). Here, attention must be paid to the fact that the helix radius depends

on the reference point which is chosen to lie on the helix surface. In our case this is the C atom in the O–C–N peptide plane. If the carbon C^{α} atom is chosen instead, we find the values given in the column with the header '$\rho_{C^{\alpha}}$'.

The parameters we find for the right-handed α-helix are very close to those given by Barlow and Thornton, who compare different standard definitions with average values computed from a set of 291 helices in 'real' proteins. The parameters listed in the above reference are in the intervals $0.23 \leq \rho \leq 0.24$, $3.54 \leq \tau \leq 3.67$ and $0.52 \leq p \leq 0.55$, respectively, using our notation and units (ρ and p in nanometres). In case of the 3_{10}-helix the spread of the parameters given by Barlow and Thornton is $0.18 \leq \rho \leq 0.20$, $3.0 \leq \tau \leq 3.2$ and $0.58 \leq p \leq 0.60$. Parameters for π-helices are not listed. It should be noted that the orientational distance takes well distinguishable values for the different secondary-structure motifs, but left- and right-handed motifs cannot be distinguished by this parameter.

3.3. Proteins in different fold classes

In the following we will show the results of *ScrewFit* for proteins which fall into the four main fold classes according to the *SCOP* scheme (Murzin *et al.*, 1995).

(i) Carbonmonoxy-myoglobin (PDB code 1a6g), which belongs to the 'all-α' class.

(ii) The protease inhibitor ecotin (PDB code 1ecy), which belongs to the 'all-β' class.

(iii) Triose phosphate isomerase from chicken muscle (PDB code 1tim), which belongs to the 'α/β' class. Proteins falling into this class consist mainly of parallel β-sheets separated by α-helices.

(iv) Hen egg-white lysozyme (PDB code 193l), which falls into the '$\alpha + \beta$' class. Proteins of this type contain mainly antiparallel β-sheets and separated regions containing α-helices.

The latter application is postponed to the next section, where we consider the structural changes of lysozyme under pressure. In this context the *ScrewFit* parameters will also be discussed in more detail. Here, we give only an impression of the results compared with *DSSP*.

Figs. 2, 3 and 4 show the comparison of the first three proteins in the list given above. In each figure we give the local orientational distance Δ_{Ω}, the local helix radius ρ and the straightness parameter σ. All calculations have been performed on the basis of the respective entries in the PDB. The vertical stripes correspond to the secondary-structure motifs found using the *DSSP* method of Kabsch and Sander, which is based on hydrogen-bonding criteria and is widely used for the determination of secondary-structure elements in proteins (Kabsch & Sander, 1983). The colouring scheme indicates α-helices in light green and β-strands in light blue.

It can be recognized that the *ScrewFit* method often leaves some ambiguity concerning the boundaries of secondary-structure elements. This is simply owing to the fact that it is sensitive to deviations from ideal geometries. This effect is particularly visible in the behaviour of the straightness parameter. Similar observations have been made by comparing the method of Barlow & Thornton (1988) using *DSSP*.

3.4. Lysozyme under hydrostatic pressure

In the following, we apply our method to visualize the structural changes in lysozyme arising from the application of external pressure. For this purpose, we consider protein structures which have been obtained from X-ray crystallography and from NMR measurements. The X-ray structures are taken from entries 193l and 3lym of the PDB, which

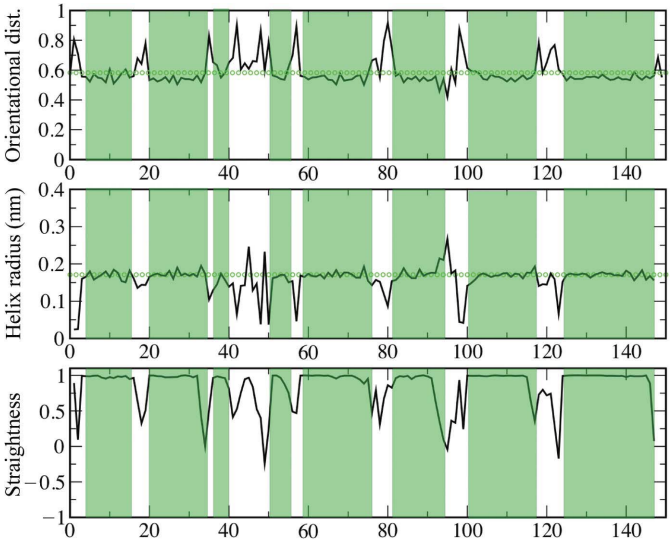

Figure 2
ScrewFit description of the main chain of carbonmonoxy-myoglobin (PDB code 1a6g; 'all-α' in the *SCOP* scheme). The vertical green stripes indicate α-helices found by the *DSSP* method and the horizontal lines indicate the reference values given in Table 1.

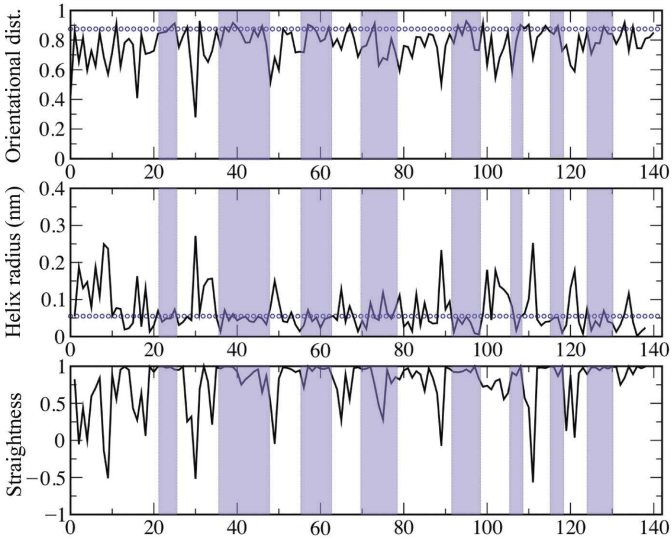

Figure 3
ScrewFit description of the main chain of the protease inhibitor ecotin (PDB code 1ecy; 'all-β' in the *SCOP* scheme). The vertical blue stripes indicate β-strands found by the *DSSP* method and the horizontal lines indicate the reference values given in Table 1.

contain the atomic coordinates of hen egg-white lysozyme at pressures of 10^5 and 10^8 Pa, respectively (Vaney *et al.*, 1996; Kundrot & Richards, 1987). The NMR structures are taken from PDB entries 1gxv and 1gxx, corresponding to pressures of 10^5 and 2×10^8 Pa, respectively (Refaee *et al.*, 2003). Fig. 5 shows the backbone of lysozyme at 10^5 Pa (blue tube) obtained from the crystal structure, together with the line joining the centres of screw motion mapping each peptide plane onto the consecutive one (red line). The centres of the screw motions have been constructed according to (22). Inspection by eye shows that that the red line passes right through the geometrical centres of the helices.

Further details can be obtained from Figs. 6 and 7, which show the same parameters as in Figs. 2, 3 and 4. In both cases the curves corresponding to the structures under pressure are given in red. The green and blue horizontal lines again correspond to the reference values for an α-helix and a β-strand, respectively, given in Table 1. Here, the vertical stripes indicate the secondary structures according to the *PDBsum* database. The latter uses the *PROMOTIF* program

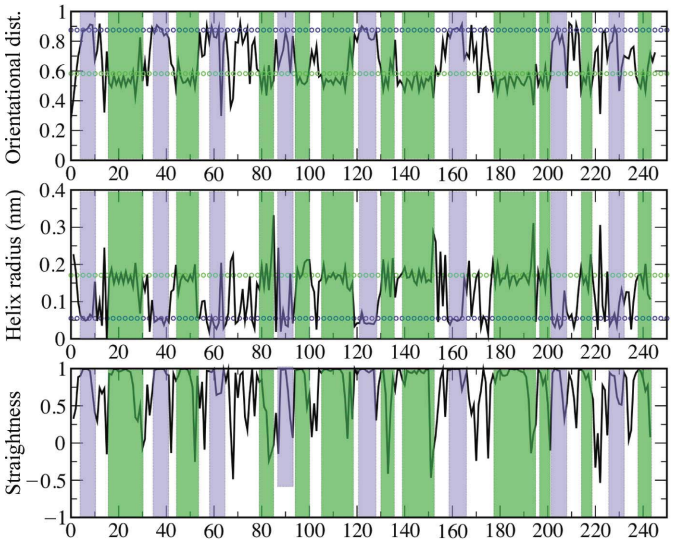

Figure 4
ScrewFit description of the main chain of triose phosphate isomerase (PDB code 1tim; 'α/β' in the *SCOP* scheme). The vertical green and blue stripes indicate α-helices and β-strands found by the *DSSP* method, respectively. The green and blue horizontal lines indicate the respective reference values from Table 1.

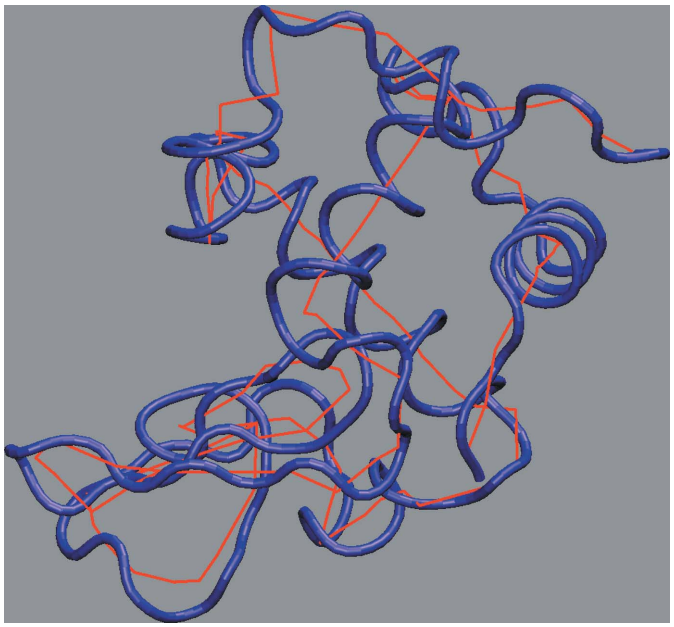

Figure 5
Minimal model for lysozyme at normal pressure. The red line joins the centres of screw motions, $X^{\perp}$, mapping each peptide plane onto the following one.

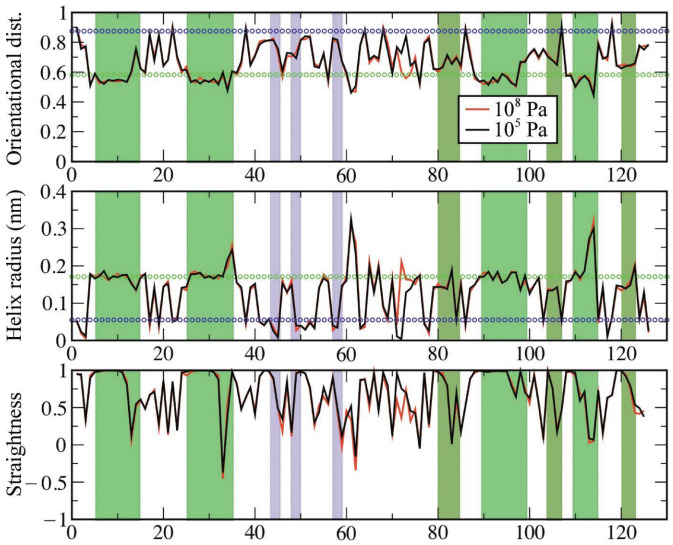

Figure 6
ScrewFit description of lysozyme for crystallographic structures at pressures of 10^5 (black line) and 10^8 Pa (red line) (PDB codes 193l and 3lym; Vaney *et al.*, 1996; Kundrot & Richards, 1987). According to the *SCOP* scheme, lysozyme falls into the '$\alpha + \beta$' class. The horizontal lines show the reference values given in Table 1 and the vertical stripes indicate the secondary structures according to the *PDBsum* database. Further explanations are given in the text

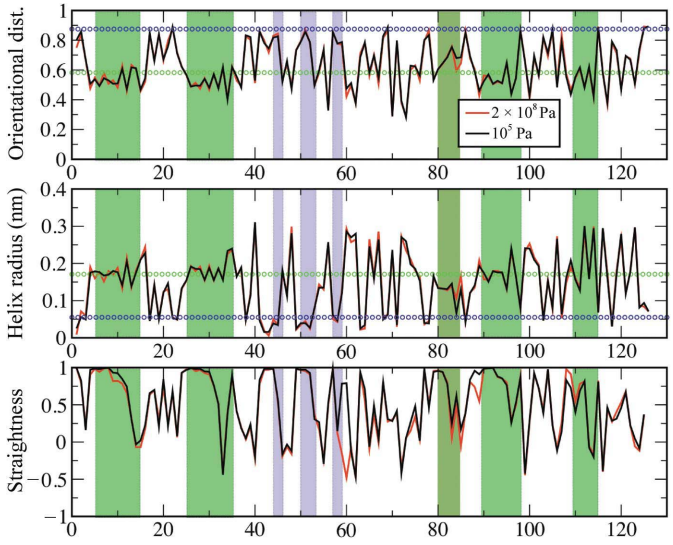

Figure 7
As Fig. 6, but for NMR structures at 10^5 (black line) and 2×10^8 Pa (red line) (Refaee *et al.*, 2003).

for secondary-structure determination (Hutchinson, 2005), which is itself based on the *DSSP* method. In addition to α-helices, we also indicate 3_{10}-helices in dark green.

For the crystal structure entries 193l and 3lym, the *PDBsum* database displays three long helices in the residue intervals 5–14, 25–36 and 89–99 and four short ones in the residue intervals 80–84, 104–107, 109–114 and 120–123. In addition, three short β-strands of two or three residues are displayed in the regions 43–45, 51–53 and 58–59, respectively. We note here that only the long helices are described in the work by Barlow & Thornton (1988). Concerning the NMR structures, the *PDBsum* database lists again the three long helices (5–14, 25–36 and 89–98), but only two short ones (80–84 and 109–114). In contrast, the short β-strands are displayed at almost the same positions as in the crystal structures (44–46, 50–53 and 58–59). As for the crystal structures, the structural motifs are found at identical positions for both pressures.

Looking first at Fig. 6, which displays the parameters corresponding to the crystal structures, shows that the orientational distance is a good measure for the rapid localization of secondary-structure elements in the amino-acid sequence of a protein. For the moment, we will only discuss the structure at ambient pressure. The analysis of the helix radius and the straightness gives more detailed information. The three long helices and also the three short β-strands are easily localized. We find that the first helix (5–14) is straight only in the region 5–11. Towards the C-terminus the straightness drops considerably and the orientational distance rises. The helix radius stays approximately constant up to about residue 15. We find that the second helix (25–36) is also deformed towards its C-terminus, but here the orientational distances stays more or less constant, whereas the helix radius and the straightness change considerably. We consider this helix to be straight in the range 25–32. Similar observations can be made for the third long helix, which we find to be straight in the range 89–96. We note here that Barlow and Thornton consider the first of the above helices as 'irregular' and the others as 'curved'; however, they use different criteria. Concerning the shorter helices, which are not considered helices by Barlow and Thornton, we confirm less well defined helices in the ranges 80–84, 104–107, 119–123 and 109–114. According to the orientational distance, the first three are 3_{10}-helices. As for the β-strands, our analysis would confirm the short strand in the range 58–59, but yield longer strands in the regions 42–46 and 50–53. It is worthwhile mentioning that the straightness parameter indicates hairpin turns between the β-strands, leading to antiparallel β-sheets.

Applying *ScrewFit* to the NMR structure of lysozyme at ambient pressure yields the following results: the three long helices indicated by *PDBsum* are retrieved and, using the orientational distance as a criterion, we confirm less well defined helices in the ranges 80–84 (3_{10}-helix) and 109–114 (α-helix). According to our analysis, the first β-strand is longer than that displayed in the *PDBsum* database (approximately 41–45).

Let us now look at the changes in secondary structure of lysozyme arising from the exertion of an external pressure. We start with the analysis of the crystallographic data of Kundrot & Richards (1987). The black line in Fig. 8 shows that the structural change obtained from the crystal structures is localized at residue 72. All parameters show a change in the

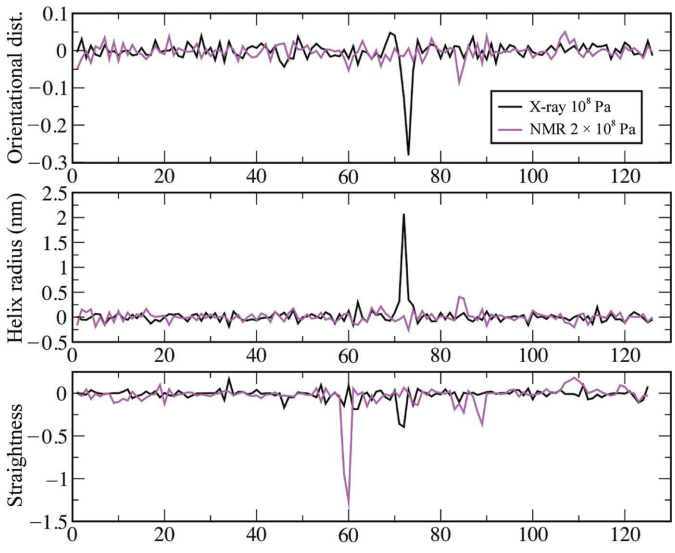

Figure 8
Differences between the parameters shown in Figs. 6 and 7 (X-ray crystallography = black line, NMR = magenta line).

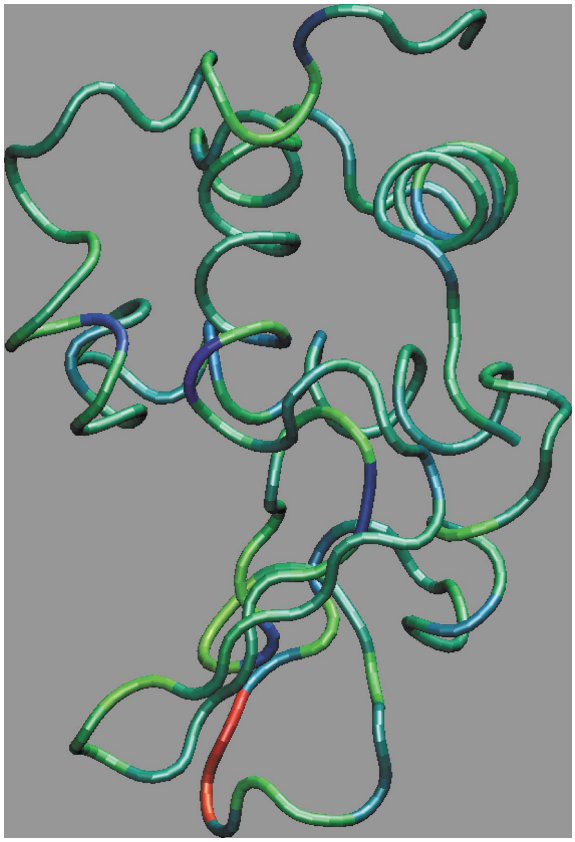

Figure 9
Change in straightness between the crystallographic structure of lysozyme at 10^8 and 10^5 Pa. The colouring scheme is chosen such that blue, green and red correspond to +2 (maximum positive change), 0 and −2 (maximum negative change), respectively.

same place. Fig. 9 shows the change in the straightness of the crystal structure of lysozyme in a tube representation, using a colouring scheme in which red corresponds to a negative change, green to no change and blue to a positive change. In their study, Kundrot and Richards perform a difference distance matrix analysis of the structural changes and report that the smallest changes are seen in helix 2 (25–36) and in the loop and β-sheet region 42–60, whereas a larger structural change is seen in region 61–87, which appears to expand. We note here that Kundrot and Richards call this region a 'loop region', not counting the short 3_{10}-helix 80–84. This observation is coherent with our observations, which show in particular a strong rise of the local helix radius at residue 72, corresponding to a swelling of the corresponding loop region.

The corresponding analysis for the NMR structures is less clear (see Fig. 7, magenta lines). Here, the orientational distance and the helix radius do not exhibit significant changes, whereas the straightness shows a strong decrease at residue 60, which is located at the very beginning of the long loop in the residue range 60–80. Refaee *et al.* (2003) report the most extensive deformations in the loop and what they call a 'β-sheet domain' (40–88), which is certainly in agreement with a very localized change in secondary structure at residue 60. However, we do not observe the considerable changes in the hairpin turns 47–49 and 54–57 seen by Refaee and coworkers. Fig. 10 shows the change in straightness for the NMR structure of lysozyme in a tube plot in which the same colouring scheme is used as in Fig. 9.

4. Conclusion

We have presented a simple method, *ScrewFit*, for the characterization of protein secondary structure that uses quaternion-based superposition fits of consecutive peptide planes in the backbone. The combined use of the quaternion fit method and Chasles' theorem allows the expression of protein secondary structure in terms of local helix parameters. The superposition method yields also an orientational distance measure for consecutive peptide planes. The latter is obtained from the 'worst' possible quaternion fit and yields a simple measure for the rapid localization of secondary-structure elements along the protein backbone. The analysis of standard motifs of protein secondary structure and of proteins belonging to different fold classes showed that all common motifs are well discriminated by the orientational distance measure and that the straightness parameter and the helix diameter are useful to characterize non-ideal secondary-structure elements, keeping a minimal set of parameters.

Using *ScrewFit* to study conformational changes in lysozyme arising from application of external pressure revealed different localized changes in the loop regions. The structural changes extracted by difference distance matrix analysis from the crystallographic data could be confirmed, however, giving a more precise description of these changes. Concerning the NMR structures, we find the essential conformational changes in a different position than the authors of the reference article, although both results agree in so far as the changes are found in the same region. Prior to these analyses, we tested that the localization of the essential secondary-structural elements found by crystallography and NMR is confirmed.

ScrewFit allows the pinpointing of secondary-structure changes precisely, which is more difficult to achieve by the standard analysis of positional differences. The reason is that the latter might indicate important structural differences in a large region, although the corresponding position differences are induced by one single localized change in the winding of the protein backbone. A point which should also be mentioned is the numerical efficiency of the quaternion-based superposition algorithm we use as a basis of our method. The superposition of two molecular structures can be performed in a few milliseconds (Kneller, 1991) and this fact has been exploited in many studies of rigid-body motions in molecular systems using the molecular-dynamics analysis package *nMoldyn* (Kneller *et al.*, 1995; Rog *et al.*, 2003). Using the method presented in this article, the characterization of the secondary structure of a protein can be performed in about 1 s on a normal PC and this efficiency could for example be used in database-oriented applications and for analyses of molecular-dynamics trajectories of proteins. In this context it is important to note that the protein backbone can be completely reconstructed from the helix parameters defined in this article. This is an interesting aspect for homology modelling. Another useful application could be the characterization of structural variability in different structural

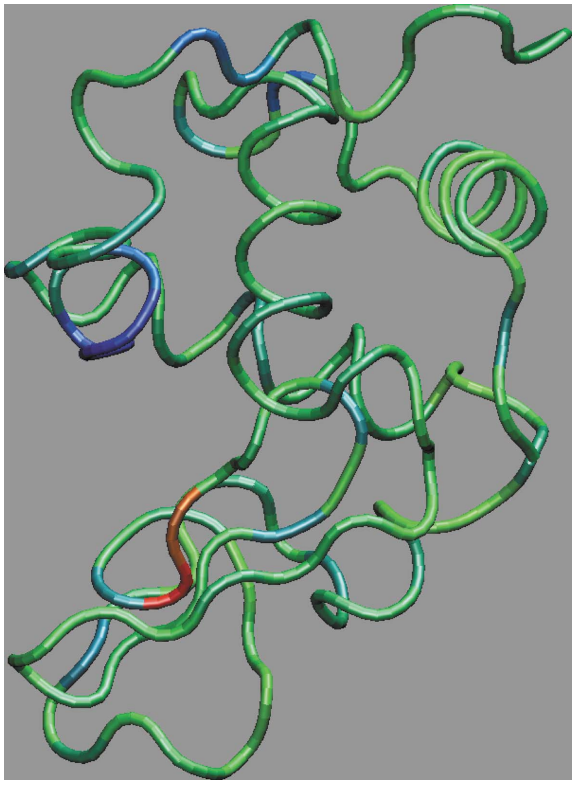

Figure 10
Change in straightness between the NMR structure of lysozyme at 2×10^8 and 10^5 Pa. The colouring scheme is the same as in Fig. 9.

models which are used to construct protein three-dimensional structures from NMR distance data.

APPENDIX *A*
Quaternions

Quaternions are hypercomplex numbers which are composed by linear superposition of one real unit element 1 and three imaginary unit elements $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$. The latter satisfy the non-commutative algebra $\mathcal{I}^2 = \mathcal{J}^2 = \mathcal{K}^2 = -1$ and $\mathcal{I}\mathcal{J} = -\mathcal{J}\mathcal{I} = \mathcal{K}$ (cycl.). An arbitrary quaternion $\mathcal{Q}$ is written as $\mathcal{Q} = q_0 1 + q_1\mathcal{I} + q_2\mathcal{J} + q_3\mathcal{K}$, where $q_j \in \mathbb{R}$ $(j = 0, \ldots, 3)$. The component q_0 is called the scalar component and $\{q_1, q_2, q_3\}$ are the vectorial components. It is useful to introduce the column vector $\mathbf{q}_v = (q_1, q_2, q_3)^T$ comprising the three vectorial components of a quaternion. Analogously to complex numbers, the length of a quaternion is defined as $\|\mathbf{Q}\| = (q_0^2 + q_1^2 + q_2^2 + q_3^2)^{1/2}$ and its conjugate is given by $\mathbf{Q}^* = q_0 1 - q_1\mathcal{I} - q_2\mathcal{J} - q_3\mathcal{K}$.

Let $\mathcal{A}$ and $\mathcal{B}$ be quaternions with components $\{a_0, a_1, a_2, a_3\}$ and $\{b_0, b_1, b_2, b_3\}$, respectively. The components of $\mathcal{C} = \mathcal{A} \pm \mathcal{B}$ are obtained by $c_j = a_j \pm b_j$ $(j = 0, \ldots, 3)$ and from the algebra of the imaginary elements one finds that the components of the product $\mathcal{C} = \mathcal{A}\mathcal{B}$ are given by

$$\begin{pmatrix} c_0 \\ \mathbf{c}_v \end{pmatrix} = \begin{pmatrix} a_0 b_0 - \mathbf{a}_v^T \cdot \mathbf{b}_v \\ a_0\mathbf{b}_v + b_0\mathbf{a}_v + \mathbf{a}_v \wedge \mathbf{b}_v \end{pmatrix},$$

where '$\wedge$' denotes a vector product. In general $\mathcal{A}\mathcal{B} \neq \mathcal{B}\mathcal{A}$. The inverse of a quaternion $\mathcal{A}$ is defined as

$$\mathcal{A}^{-1} = \frac{\mathcal{A}*}{\|\mathcal{A}\|^2}.$$

Owing to the non-commutative algebra of quaternions, in general $\mathcal{A}^{-1}\mathcal{B} \neq \mathcal{B}\mathcal{A}^{-1}$.

Similarly to complex numbers of unit length, which represent rotations in the plane, normalized quaternions represent rotations in space. Let $\mathbf{r} = (x, y, z)^T$, a column vector comprising the components of a radius vector $\vec{r}$, let $\mathcal{R} = x\mathcal{I} + y\mathcal{J} + z\mathcal{K}$ be the corresponding spatial quaternion and let $\mathcal{Q}$ be a normalized quaternion with $\|\mathcal{Q}\| = 1$. It is found that the scalar component of $\mathcal{R}' = \mathcal{Q}\mathcal{R}\mathcal{Q}^*$ also vanishes and that the vectorial components of $\mathcal{R}'$ are given by

$$\mathbf{r}' = \mathbf{D} \cdot \mathbf{r},$$

where $\mathbf{D}$ represents the rotation matrix (2). The bilinear transformation $\mathcal{R}' = \mathcal{Q}\mathcal{R}\mathcal{Q}^*$ thus represents a rotation in space.

APPENDIX *B*
Helix parameters in Chasles' theorem

Chasles' theorem can easily be proven using quaternion algebra. For this purpose we start from (18) and introduce the spatial quaternions $\mathcal{U}$ and $\mathcal{T}_\perp$, representing, respectively, the column vectors $\mathbf{u}$ and $\mathbf{t}_\perp$. Expressed in quaternions, (18) becomes

$$\mathcal{U} - \mathcal{Q}\mathcal{U}\mathcal{Q}^* = \mathcal{T}_\perp.$$

Multiplication with $\mathcal{Q}$ from the right and using the fact that $\mathcal{Q}^*\mathcal{Q} = 1$ yields

$$\mathcal{U}\mathcal{Q} - \mathcal{Q}\mathcal{U} = \mathcal{T}_\perp\mathcal{Q}.$$

Using the multiplication rule for quaternions, the above equation can be expressed in the form

$$\begin{pmatrix} -\mathbf{u}^T \cdot \mathbf{q}_v \\ q_o\mathbf{u} + \mathbf{u} \wedge \mathbf{q}_v \end{pmatrix} - \begin{pmatrix} -\mathbf{q}_v^T \cdot \mathbf{u} \\ q_o\mathbf{u} + \mathbf{q}_v \wedge \mathbf{u} \end{pmatrix} = \begin{pmatrix} -\mathbf{t}_\perp \cdot \mathbf{q}_v \\ q_0\mathbf{t}_\perp + \mathbf{t}_\perp \wedge \mathbf{q}_v \end{pmatrix}.$$

Here, we can make use of relation (11),

$$\begin{pmatrix} q_0 \\ \mathbf{q}_v \end{pmatrix} = \begin{bmatrix} \cos(\varphi/2) \\ \sin(\varphi/2)\mathbf{n} \end{bmatrix},$$

from which we conclude that $\mathbf{t}_\perp \cdot \mathbf{q}_v = 0$, since $\mathbf{t}_\perp \perp \mathbf{n}$. We are thus left with the vector equation

$$\begin{pmatrix} 0 \\ 2\mathbf{u} \wedge \mathbf{q}_v \end{pmatrix} = \begin{pmatrix} 0 \\ q_0\mathbf{t}_\perp + \mathbf{t}_\perp \wedge \mathbf{q}_v \end{pmatrix},$$

which can be reduced to

$$\mathbf{n} \wedge \mathbf{u} = \tfrac{1}{2}[-\cot(\varphi/2)\mathbf{t}_\perp + \mathbf{n} \wedge \mathbf{t}_\perp] \qquad (29)$$

if $\varphi \neq 2k\pi$ $(k \in Z)$. Now one can apply on both sides a vectorial multiplication with $\mathbf{n}$, using the fact that $\mathbf{n} \wedge (\mathbf{n} \wedge \mathbf{a}) = -\mathbf{a}_\perp$ for an arbitrary column vector $\mathbf{a}$. This yields

$$\mathbf{u}_\perp = \tfrac{1}{2}[\mathbf{t}_\perp + \cot(\varphi/2)\mathbf{n} \wedge \mathbf{t}],$$

if one uses that $\mathbf{n} \wedge (\mathbf{n} \wedge \mathbf{t}_\perp) = -\mathbf{t}_\perp$ and that $\mathbf{n} \wedge \mathbf{t}_\perp = \mathbf{n} \wedge \mathbf{t}$. Relation (20) is thus proven. The general solution of (29) obviously has the form

$$\mathbf{u}(\lambda) = \mathbf{u}_\perp + \lambda\mathbf{n}, \quad \lambda \in \mathbb{R},$$

which shows that $\mathbf{u}_\perp(\lambda)$ is the solution of minimum length.

All figures containing molecular graphics were generated using the *VMD* code for molecular-dynamics simulation and visualization of biomolecules (Humphrey *et al.*, 1996). The screw-motion calculations were performed with modules from the *MMTK* package (Hinsen, 2000).

References

Altmann, S. (1986). *Rotations, Quaternions and Double Groups.* Oxford: Clarendon Press.
Barlow, D. & Thornton, J. (1988). *J. Mol. Biol.* **201**, 601–619.
Chasles, M. (1830). *Bull. Sci. Math.* **14**, 321–326.
Chasles, M. (1870). *Rapport sur les Progès de la Géométrie en France*, p. 77. Paris: Imprimérie Nationale.
Frishman, D. & Argos, P. (1995). *Proteins*, **23**, 566–579.
Hinsen, K. (2000). *J. Comput. Chem.* **21**, 79–85.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Hutchinson, G., (2005). *Promotif* v.1.0/v.2.0. http://www.biochem.ucl.ac.uk/bsm/promotif/promotif.html.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kearsley, S. (1989). *Acta Cryst.* A**45**, 208–210.
Kneller, G. (1991). *Mol. Sim.* **7**, 113–119.
Kneller, G. (2005). *J. Comput. Chem.* **26**, 1660–1662.
Kneller, G., Keiner, V., Kneller, M. & Schiller, M. (1995). *Comput. Phys. Commun.* **91**, 191–214.
Kundrot, C. & Richards, F. (1987). *J. Mol. Biol.* **193**, 157–170.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Quine, J. (1999). *J. Mol. Struct. (Theochem)*, **460**, 53–66.
Refaee, M., Tezuka, T., Akasaka, K. & Williamson, M. (2003). *J. Mol. Biol.* **327**, 857–865.
Richards, F. & Kundrot, C. (1988). *Proteins*, **3**, 71–84.
Rog, T., Murzyn, K., Hinsen, K. & Kneller, G. (2003). *J. Comput. Chem.* **24**, 657–667.
Selig, J. (1996). *Geometrical Methods in Robotics.* Berlin: Springer.
Sklenar, H., Etchebest, C. & Lavery, R. (1989). *Proteins*, **6**, 46–60.
Stryer, L. (1988). *Biochemistry.* New York: W. H. Freeman & Co.
Taylor, W. (2001). *J. Mol. Biol.* **310**, 1135–1150.
Thomas, D. J. (1994). *J. Mol. Graph.* **12**, 146–152.
Vaney, M., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.