Journal of Molecular Graphics and Modelling 38 (2012) 256–278

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

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

Quaternion maps of global protein structure

Andrew J. Hanson[a,*], Sidharth Thakur[b]

[a] School of Informatics and Computing, Lindley Hall 215, Indiana University, Bloomington, IN 47405, USA
[b] Renaissance Computing Institute, Suite 540: 100 Europa Drive, Chapel Hill, NC 27517, USA

ARTICLE INFO

Article history:
Accepted 14 June 2012
Available online 23 June 2012

Keywords:
Visualization
Proteins
Amino acids
Orientation frames
Quaternions
Quaternion maps

ABSTRACT

The geometric structures of proteins are vital to the understanding of biochemical interactions. However, there is much yet to be understood about the spatial arrangements of the chains of amino acids making up any given protein. In particular, while conventional analysis tools like the Ramachandran plot supply some insight into the local relative orientation of pairs of amino acid residues, they provide little information about the global relative orientations of large groups of residues. We apply quaternion maps to families of coordinate frames defined naturally by amino acid residue structures as a way to expose global spatial relationships among residues within proteins. The resulting visualizations enable comparisons of absolute orientations as well as relative orientations, and thus generalize the framework of the Ramachandran plot. There are a variety of possible quaternion frames and visual representation strategies that can be chosen, and very complex quaternion maps can result. Just as Ramachandran plots are useful for addressing particular questions and not others, quaternion tools have characteristic domains of relevance. In particular, quaternion maps show great potential for answering specific questions about global residue alignment in crystallographic data and statistical orientation properties in Nuclear Magnetic Resonance (NMR) data that are very difficult to treat by other methods.

1. Introduction

We explore a family of global visualization methods for exploiting quaternion maps of intrinsic protein orientation frames. The advantage of quaternion maps is that a single quaternion point embodies the full three degree-of-freedom transformation from the identity frame triad in three dimensions (3D) to an arbitrary frame triad; therefore, a quaternion frame representation is much simpler than the usual frame representation using a triple of orthogonal 3D vectors, and simultaneously it is much richer than the Ramachandran plot, which in addition to having only two degrees-of-freedom, can only represent the relative orientations of immediately neighboring residues. Quaternions naturally expose global similarities among all residues in a protein complex, no matter how near or how distant, and extending across component proteins in multi-part structures.

Our canonical methods for visually representing quaternion values as geometric points in space superficially resemble geographic maps of the world globe, but the distinction is profound: while similar problems are addressed by the relation between a flat globe map such as a Mercator projection and the actual spherical surface of the Earth, the Mercator-to-Globe relationship is identical in dimension and produces identical local representations in a sufficiently small neighborhood (your town's map is flat for all practical purposes). There is no local correspondence between a quaternion and a frame triad: a quaternion is a point in four Euclidean dimensions, constrained to move inside a particular three-degree-of-freedom spherical space, while a frame triad contains nine separate components constrained to have the properties of a 3×3 orthogonal matrix, a completely distinct representation of the three-degree-of-freedom 3D orientation system. The correspondence also has a deep mathematical context: one of the greatest geometric achievements of the 19th century was the discovery of the quadratic form, based on quaternions, that embodies an exact map from a quaternion point to the 9 elements of an arbitrary 3×3 orthogonal rotation matrix (an arbitrary 3D frame triad in our context), together with the reverse two-fold ambiguous mapping from any such frame matrix to a quaternion point.

Here we investigate the details of these mappings as they can be applied to reveal properties of the spatial orientations of protein systems. Section 2 reviews previous work in this area. Section 3 outlines the mathematical and geometrical properties of quaternions that we will be exploiting, with additional details provided in the Appendices. Two classes of quaternion visualization methods are provided, one based on a visual geometric context ("geometric view"), the other based on parallel coordinates and some innovative quaternion-driven variants ("coordinate view"). Section 4 provides numerous intuition-building examples of quaternion frame methods applied first to ideal mathematical curves, then to idealized

* Corresponding author. Tel.: +1 812 855 5855.
E-mail addresses: hansona@indiana.edu (A.J. Hanson), sthakur@renci.org (S. Thakur).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.06.004

spline curves used traditionally to represent a high-level protein structure. Finally, in Section 5, we illustrate applications to discrete frames given by the atomic positions of residue components in a protein's PDB file. Section 6 expands our scope to a variety of protein data domains and applications, including in particular a treatment of the orientation variations present in the statistical distributions of NMR data. Section 7 summarizes the spectrum of tools that can be applied to studies of quaternion maps, and a lengthy appendix is devoted to a pedagogical study of the relation between Ramachandran plots and our quaternion maps. In summary, quaternion maps have the potential to expose novel properties and features of protein geometry, with particular applicability to questions of global overall structure.

2. Related work

While quaternions have been employed extensively to encode molecular orientations (see [1–5]), and have also been applied to RNA (see [6]), applications of quaternions to protein structures have been limited in scope (see, for example, [7–14]). The most widely used approach to analyzing orientations of amino acid residues is the classic work of Ramachandran [15,16], which encodes only local information about orientation angles, although alternative orientation visualization methods have been proposed, e.g., by Bojovic et al. [17].

A number of interesting approximations to the Ramachandran information, along with techniques that exploit quaternion derivatives, have been explored by R. Hanson et al. [18]. Part of the latter work was in fact motivated by an unpublished version of the current manuscript; the *jmol* molecular visualization system now includes the QUATERNION command implementing a number of the basic quaternion mapping functions we describe here. Other treatments, such as Morris et al. [19], use both local and global structures to ascertain the stereochemical nature of proteins, but their visualizations of protein stereochemistry are limited to two-dimensional plots and histograms. Our treatment here is somewhat complementary to these, focusing on visualizing global residue orientation properties directly in quaternion space (the "quaternion Gauss map" [20–23]).

3. Introduction to quaternion maps

This section introduces orientation frames in quaternion form, the geometric view of quaternions, and the coordinate view of quaternions, along with the extension of single quaternion point displays to the display of a series of quaternion points.

3.1. Quaternion orientation frames

An *Orientation Frame* **F** can be specified as a triple of mutually orthogonal normalized three-vectors, where the identity frame consists of the three columns composed of the x-axis, the y-axis, and the z-axis. Any frame whatsoever can, by a theorem of Euler, be expressed as a rotation $\mathbf{R}(\theta, \hat{\mathbf{n}})$ that acts on the identity frame and rotates it about a fixed direction $\hat{\mathbf{n}}$ by some angle θ, where $\hat{\mathbf{n}}$ is the unique real eigenvector of $\mathbf{R}(\theta, \hat{\mathbf{n}})$. The columns of the matrix **R** are exactly the three vectors describing the corresponding frame **F**.

Rotation matrices and the actions of rotations in three dimensions, and hence orientation frames, can alternatively be represented by *unit-length quaternions* (see, e.g., [23]). Just as a unit-length complex number $\cos\theta + i\sin\theta = \exp(i\theta)$ with $i^2 = -1$ can be represented by a pair of real numbers (x, y) satisfying $x^2 + y^2 = 1$, a unit-length quaternion $(q_0 + \mathbf{i}q_x + \mathbf{j}q_y + \mathbf{k}q_z) = \exp(\mathbf{I}\cdot\hat{\mathbf{n}}(\theta/2))$ with $\mathbf{I} = (i, j, k)$ and $i^2 = j^2 = k^2 = ijk = -1$ can be represented as a quadruple of real numbers

$$q(\theta, \hat{\mathbf{n}}) = (q_0, q_x, q_y, q_z) = (w, \mathbf{x}) = \left(\cos\left(\frac{\theta}{2}\right), \hat{\mathbf{n}}\sin\left(\frac{\theta}{2}\right)\right) \tag{1}$$

where it is sometimes convenient to define $w = q_0$ and $\mathbf{x} = (x, y, z) = (q_x, q_y, q_z)$. Here the rotation axis $\hat{\mathbf{n}}$ and the angle θ correspond precisely to those introduced already in $\mathbf{R}(\theta, \hat{\mathbf{n}})$; it is easy to verify that this parameterization has unit length, $q\cdot q = q_0^2 + q_x^2 + q_y^2 + q_z^2 = 1$, and has only the obligatory three free rotation parameters since $\hat{\mathbf{n}}\cdot\hat{\mathbf{n}} = 1$ as well. The solutions of $q\cdot q = 1$ (which define the three-dimensional topological space of unit quaternions) are known as $\mathbf{S}^3$, or the *three-sphere*.

Quaternions as represented in Eq. (1) have some additional properties of particular interest to us here:

- There exists a quadratic formula that defines a two-to-one mapping from a quaternion q to a frame represented as a 3×3 rotation matrix **R**, and, given a rotation matrix, one can find the two unique corresponding diametrically opposed quaternions (see Appendix A).
- The identity frame corresponds both to the quaternion $q = (1, 0, 0, 0)$ and to $q = (-1, 0, 0, 0)$.
- If we require two quaternions to multiply together using the following order-dependent (non-commutative) rule originally discovered by Hamilton,

$$Q = q_1 \star q_2 = (w_1 w_2 - \mathbf{x}_1\cdot\mathbf{x}_2,\ w_1\mathbf{x}_2 + w_2\mathbf{x}_1 + \mathbf{x}_1\times\mathbf{x}_2)$$

  where $\star$ is quaternion multiplication, the resulting quaternion Q *remains embedded* in $\mathbf{S}^3$ and generates the composite 3×3 rotation matrix $\mathbf{R} = \mathbf{R}_1\cdot\mathbf{R}_2$. We reiterate that the order matters: neither quaternions nor 3D rotation matrices commute in general.
- The inverse of $q = q(\theta, \hat{\mathbf{n}})$ is just $q^{-1} = q(\theta, -\hat{\mathbf{n}}) = q(-\theta, \hat{\mathbf{n}})$, and corresponds to the inverse 3D rotation matrix.
- There exists a *quaternion distance formula*, $d_{12} = \theta_{12} = 2\cos^{-1}(q_1\cdot q_2)$, that gives a precise and rigorous definition of the similarity between frames (see Appendix B). This corresponds intuitively to a great circle or geodesic minimal-length arc connecting two points on an ordinary sphere. Smooth spline-like curves, based on the properties of the distance formula and embedded in the quaternion sphere $\mathbf{S}^3$, can be constructed that smoothly connect sequences of quaternion points [28].

3.2. Visualizing quaternions as geometry

A simple example of a frame **F** resulting from applying a rotation about the $\hat{\mathbf{z}}$ direction to the identity frame is shown in Fig. 1(a). Our next task is to relate this frame to its quaternion representation and to convert the standard 3D display of this frame to a quaternion display. In this subsection, we will explore explicit geometric views of the frame quaternion, and in the following subsection, we will examine alternative coordinate views.

For a positive (counterclockwise) rotation about the z-axis, the matrix **F** becomes

$$\mathbf{F} = \mathbf{R}(\theta, \hat{\mathbf{z}}) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

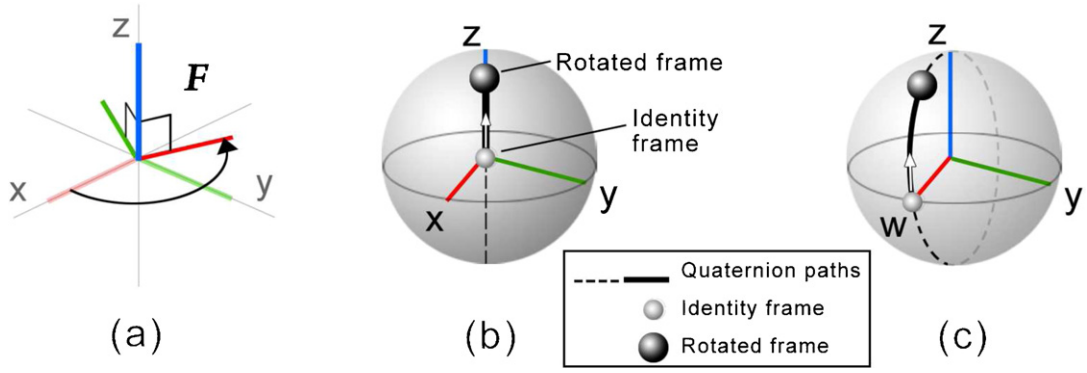


Fig. 1. (a) A simple frame **F** obtained by rotating the identity frame by a 160° angle about the direction $\hat{\mathbf{z}}$. The smaller white spheres in the schematic diagrams (b) and (c) show the location of the identity frame and the larger black spheres show the quaternion points representing the outcome of the rotation. The thick black lines show the quaternion path starting at the identity, and the dashed lines show the path of a continued z-axis rotation. (b) The canonical (x, y, z) projection, which produces a cyclic line up and down the z-axis. (c) The (w, y, z) projection, which produces a circle in the (w, z) plane.

and the columns are easily seen to be the components of **F**. Since we know that the fixed rotation axis is $\hat{\mathbf{n}} = (0, 0, 1)$, we can also write down the quaternion from Eq. (1) as

$$q_{\mathbf{F}}(\theta, \hat{\mathbf{z}}) = \left(\cos\left(\frac{\theta}{2}\right),\ 0,\ 0,\ \sin\left(\frac{\theta}{2}\right)\right).$$

How do we use this information to make a geometric view of $q_{\mathbf{F}}$? We have already remarked that any unit-length quaternion four-vector q corresponds to a point on the three-manifold $\mathbf{S}^3$, a three-sphere embedded in four Euclidean dimensions. However, there is a simple trick that allows us to make an accurate 3D picture of this complicated 4D object. Since our quaternions have unit length, the fourth component w is redundant and is just the solution of the equation $w = \pm\sqrt{1 - x^2 - y^2 - z^2}$. Hence, up to a sign, all the orientation-frame information embodied in a quaternion can be represented by a point (x, y, z) in 3D Euclidean space. We can generally choose the positive sign without losing critical information, and we can plot the three-vector component of any quaternion as a point $\boxed{\hat{\mathbf{n}}\sin(\theta/2)}$ inside a unit-radius solid sphere; this solid sphere is technically one of the two "hemispheres" of $\mathbf{S}^3$. We rarely need to dwell on this because, in practice, we study sequences of linked frames that do not frequently cross between hemispheres; nevertheless, it is important to be aware of the possibility when it happens, and to be prepared, e.g., to use different colors to encode in which hemisphere a point lies. Another variant of the geometric view is to choose alternative projections, picking, say, $x = \pm\sqrt{1 - w^2 - y^2 - z^2}$ instead of w as the "extra" variable, and plotting the point (w, y, z) inside a distinct unit-radius solid ball. These two choices are represented in quaternion coordinates in Fig. 1(b,c) for rotations about the $\hat{\mathbf{z}}$ axis leading from the identity quaternion to $q_{\mathbf{F}}$, the quaternion representation of our sample frame **F**.

3.3. *Visualizing quaternions as coordinates*

In our experience, most situations involving frame comparisons are most effectively represented using the geometric view of quaternion coordinates. Nevertheless, in some cases one may prefer a very explicit (if less visually intuitive) representation showing a list of quaternion coordinate values. The conventional representation of this type is the *parallel coordinate representation* [24]. This representation in our case would consist of taking a 4D quaternion vector representing an orientation frame (in some fixed, arbitrary order in 4D such as (q_0, q_x, q_y, q_z)) and drawing three lines connecting a graph of those four numbers, giving a display for one point like that in Fig. 2(a) corresponding exactly to Fig. 1(b,c).

Another widely used coordinate view for multidimensional data is the *star plot* (see, e.g., Chambers et al. [25] or Fanea et al. [26]). In this approach, the real line in Fig. 2(a) is essentially deformed to a

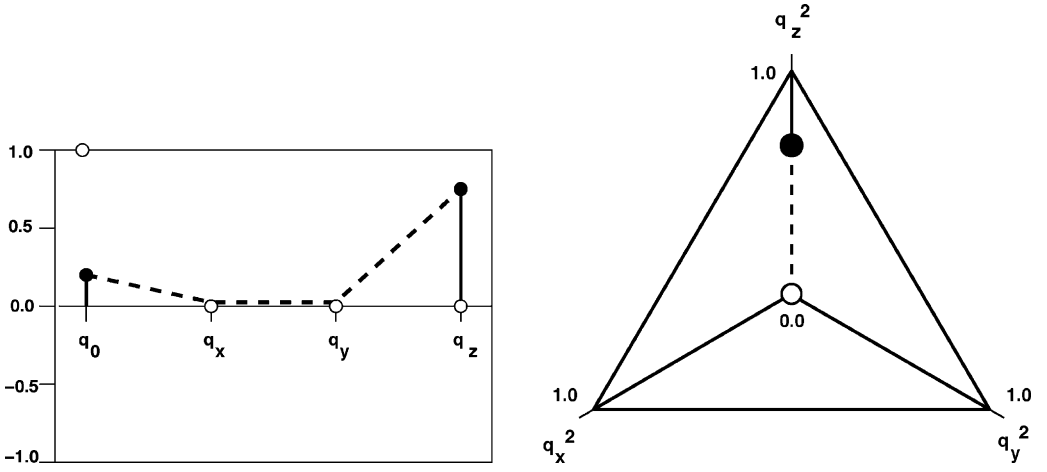


Fig. 2. (a) A classic parallel coordinate map for a single quaternion point such as the frame in Fig. 1. The four quaternion coordinates are represented in this case by the four values placed side by side and connected by three line segments to denote a single point. (b) The quaternion star-squared map (see Eq. (2)), showing the quaternion frame of **F** as a line from the x^2 and y^2 points at the origin to the z^2-value on the vertical axis. White dots denote zeroes or identity-frame values.

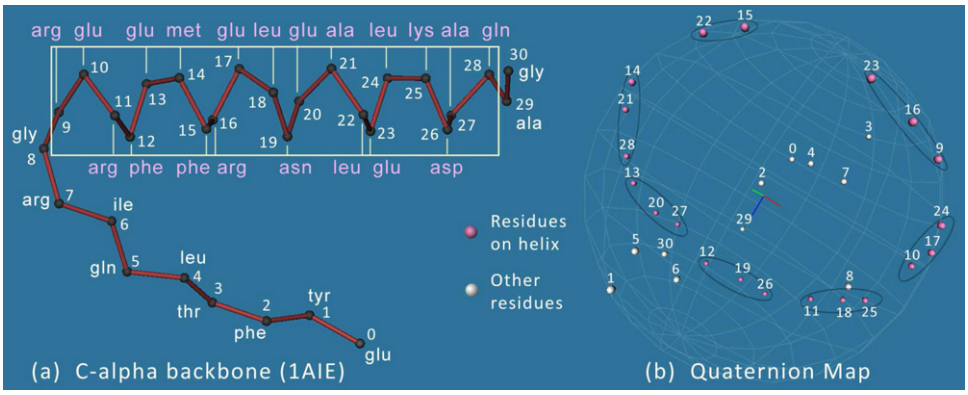


Fig. 3. (a) The PDB file geometry of 1AIE containing an alpha helix. (b) The quaternionic geometric view of the 1AIE quaternion frame coordinates in the *wyz* projection.

point and the graph connecting the four coordinate values becomes a piecewise-linear circle bounded by a diamond-shaped polygon. However, because quaternions have the special property of unit length, we again can pick three independent coordinates instead of four constrained coordinates and construct a three-axis star plot instead of a four-axis star plot.

There are several variants to the star plot. The four-axis quaternion star plot is sufficiently similar to the parallel coordinate plot that we will omit detailed discussion. The three-axis star plot is based on the 3×2 projection matrix

$$\sigma = \begin{bmatrix} -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}. \tag{2}$$

From this matrix we can construct several variants:

- Plot the triangle formed by $(\sigma \cdot (x, 0, 0)^t, \sigma \cdot (0, y, 0)^t, \sigma \cdot (0, 0, z)^t)$. This has the advantage of placing the identity frame uniquely at the origin, but negative values of z appear in the same region as positive values of x and y.
- Modifying the above by displacing the *xyz* coordinates to be always positive, i.e., $(x+1, 0, 0)^t$, etc., effectively makes the graphing areas unique. However, the identity frame is now an odd finite triangle.
- A novel variation is to map the identity frame uniquely to the origin by using the absolute values or the squares of the coordinates in the map. Although we lose some specific information, we prefer the star plot of the *squared* values, that is $(x^2, 0, 0)^t$, etc., which shows clearly how close the quaternion value is to the identity frame (the center of the star plot).

In Fig. 2(b) we show the star-squared quaternion plot corresponding to Fig. 1(b,c). For our simple z-axis rotation example, all one sees is a line to a point on the vertical axis. More general frames would lie inside an equilateral triangle, with frames near the identity frame converging on the center. Remark: one could also produce star-like plots with *single* 2D points for each frame instead of a triangle, such as $\sigma \cdot (x^2, y^2, z^2)^t$, but any equal-weighted value of (x, y, z) is confused with the identity.

3.4. Collections of frames

A significant property of the quaternion views just described is that they provide a visual image of an orientation frame as a single point (or graph) in space, the *quaternion map*. We now show how quaternion maps can be used to expose the absolute similarity of two 3D orientation frames (arbitrarily separated in 3D distance) using the proximity of the two quaternion points in the plot (see Appendix B). Selected groups of dozens of orientation frames occurring at widely different spatial positions may correspond to quaternion points falling close to one another or in a revealing pattern in the quaternion map. Dissimilarities can similarly be exposed. While the geometric view has the most powerful tools for exposing global similarities, the parallel coordinate or star plot approaches to representing frames with the coordinate view can also suggest interesting relationships among frames.

For a *collection of frames* (in our case, a set of frames corresponding to a sequence of residues in a protein), each frame is then represented as a distinct point or graph of some sort, and the *ordered sequence* of frames can be represented by a collection of these. A special technique is typically used for sequences of quaternion frames to enforce continuity of the quaternion value: since any frame can be represented by *either* q or $-q$, we must eventually choose one. We therefore compare the inner product $q_k \cdot q_{k+1}$ for each neighboring pair of frames $(k, k+1)$, and replace $q_{k+1} \to -q_{k+1}$ if the inner product is negative.

Our choices of representations that *embody the intrinsic quaternion distances* include the following:

- **S^3 Map.** Using the projection directly from the 4D quaternion value in S^3 to a 3D subspace such as *xyz* produces a spherically deformed map of the actual quaternion distances (like looking at a country on a globe of the Earth from an oblique angle). However, the deformation is completely predictable, and distances for pairs near the center, for example, are reliable. Interactive 4D rotations (see [23,27]) can place pairs anywhere one would like in the projection. The metrically most accurate distance in the projection is found by transforming the scene so one of the desired pair of quaternion frames is at the origin in the *xyz* projection.

  As a matter of practice, the curves connecting collections of quaternion points in the spherical projection are typically drawn as geodesics (shortest-distance paths constrained to the three-sphere), though for simplicity one may also draw them as piecewise linear paths. The advantage of this representation is that no matter how distant one object is from another in 3D space or along the sequence, objects that have similar orientation frames can always be forced to have nearby quaternion points.
- **Parallel Coordinate Map.** A typical parallel coordinate plot for a collection has all the points superimposed on a single 2D plot. Our case is different from the usual case because we have the *additional* quaternion distance information available, which we can use to *displace* each set of plotted 4D coordinates from its neighbors in a meaningful way. That is, we take each individual parallel coordinate plot and displace it in the perpendicular direction by the value of the quaternion distance to its next neighbor. This approach has the advantage that *all* absolute orientation differences from neighboring frames around the curve are represented

as completely accurate Euclidean distances. It has the disadvantage that it is hard to get an intuitive feeling of whether spatially distant objects (far apart in space and/or the parallel coordinate plot) have similar quaternion frame values or not.

- **Star Square Map.** Sequences of the star maps that we have defined using the projection Eq. (2) can also be displaced perpendicularly relative to each neighbor by the amount of the quaternion distance. Again, this gives a metrically accurate neighboring frame-distance representation for large numbers of frames, and exposes patterns that have visual advantages similar to the geometric projection, but relating distant frames to one another is a challenge.

These methods are illustrated for a generic sequence of frames taken from the 1AIE PDB data in Figs. 3–6.

Summary. This completes our treatment of some of the ways that, once we have *single* instances of quaternion frames, we can start to keep track of *sequences* of quaternion frames in various contexts. Each method has specific domains of utility. Our own preference is for representations with clear geometric properties as opposed to coordinate-value properties, and thus we will for the most part choose the geometric quaternion point projections in *xyz* or *wyz* coordinates.

4. Studies in quaternion frame maps

We now explore some specific examples of frame maps for idealized mathematical objects that correspond closely to the behavior of real protein data, providing additional intuitive grounding. The following section will show a parallel sequence of examples taken directly from PDB file data.

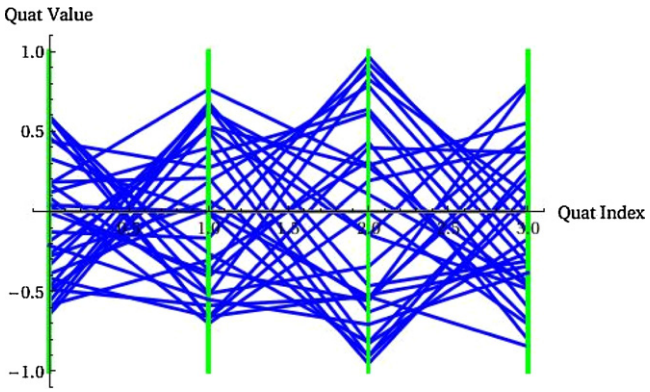


Fig. 4. Standard parallel coordinate plot of the 1AIE standard quaternion frame 4-vectors, with no means of distinguishing neighboring residues.

4.1. Alpha-helix model: quaternion frames of idealized curves

We now turn to an elementary application of quaternion Frenet–Serret frames [21] to the study of helical curves, which correspond to alpha-helices in proteins. Due to the double-valued nature of the relation between quaternions and rotations, two full turns of the helix correspond to exactly one closed circuit in quaternion space. The quaternion map in this case is a circular closed loop

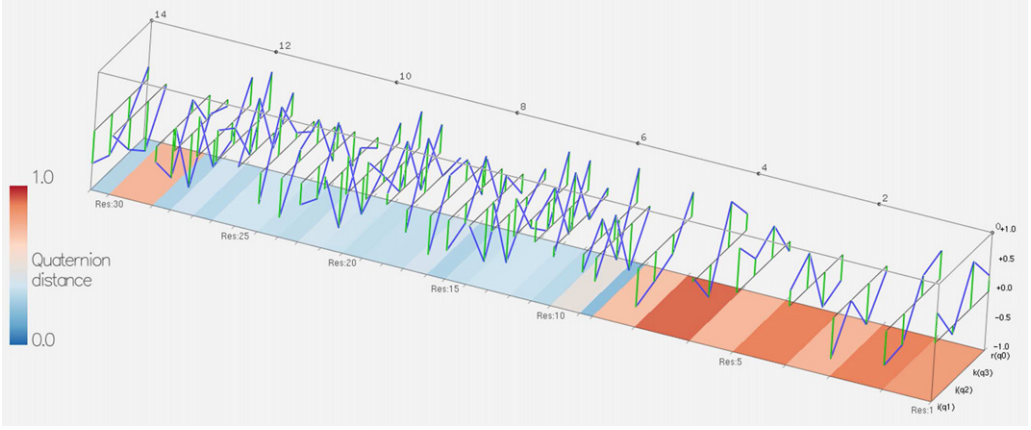


Fig. 5. Parallel coordinate plot of the 1AIE standard quaternion frame 4-vectors, spaced by quaternion distance.

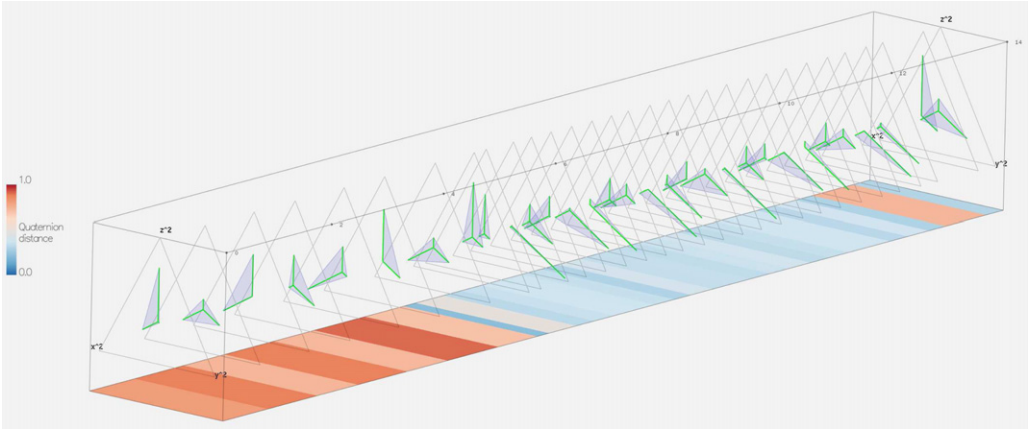


Fig. 6. Star-squared parallel coordinate plot of the 1AIE quaternion frame sequence, spaced by quaternion distance.

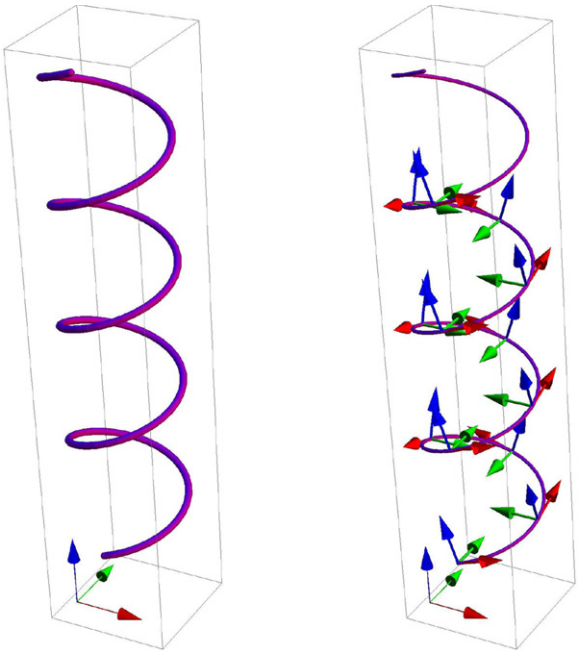


Fig. 7. (a) A helix defined by the parametric equation $(r\cos(t), r\sin(t), pt)$. (b) A set of frames on the helical curve defined by the Frenet–Serret equation. Note the relation of the identity frame at bottom left to the first actual helix frame.

that has an elliptical projection into the (x, y, z) coordinates, determined by the axis direction of the helix and its pitch. Fig. 7 shows an ideal mathematical helix and a sampling of the continuous Frenet–Serret frames determined by the local curve derivatives; Fig. 8 presents the corresponding *xyz* and *wyz* quaternion maps of the orientation frames in Fig. 7. Note that the Frenet–Serret frame may not be suitable for certain classes of curves; if there is a straight section or an inflection point (typical of cubic curves, for example), the second derivative vanishes and the Frenet–Serret frame becomes undefined.

A quick outline of how one actually does a quaternion calculation for a helix may prove useful for understanding the quaternion frames of an alpha helix structure. We start with the equations of a helix of radius r and pitch p, along with its first and second derivatives:

$$\mathbf{x}(t) = (a\cos t,\ b\sin t,\ pt)$$
$$\mathbf{x}'(t) = (-a\sin t,\ b\cos t,\ p)$$
$$\mathbf{x}''(t) = (-a\cos t,\ -b\sin t,\ 0).$$

We take $a = b = r$ to produce a circular helix, and $a \gg b$ to make a flattened elliptical helix. We use the value of the tangent (first derivative) to determine the direction of the first frame axis, which we label **X**. Typically, the next frame axis direction is computed from the cross-product $\mathbf{x}'(t) \times \mathbf{x}''(t)$, whose direction we label **Z**; then the remaining frame axis direction is $\mathbf{Y} = \mathbf{Z} \times \mathbf{X}$. Normalizing to unit length, we obtain the result for the frame triad of vectors for any point t on the helix:

$$\begin{aligned}
\mathbf{X}(t) &= \left(-\frac{r\sin(t)}{\sqrt{p^2+r^2}}, \frac{r\cos(t)}{\sqrt{p^2+r^2}}, \frac{p}{\sqrt{p^2+r^2}}\right) \\
\mathbf{Y}(t) &= (-\cos(t),\ -\sin(t),\ 0) \\
\mathbf{Z}(t) &= \left(\frac{p\sin(t)}{\sqrt{p^2+r^2}}, -\frac{p\cos(t)}{\sqrt{p^2+r^2}}, \frac{r}{\sqrt{p^2+r^2}}\right).
\end{aligned} \tag{3}$$

Do not forget that these are the *column* vectors for the frame matrix, not the *row* vectors.

The quaternion frame can then be computed as a rotation about the z axis acting on the initial frame at $t = 0$, which reduces after a bit of algebra to the form

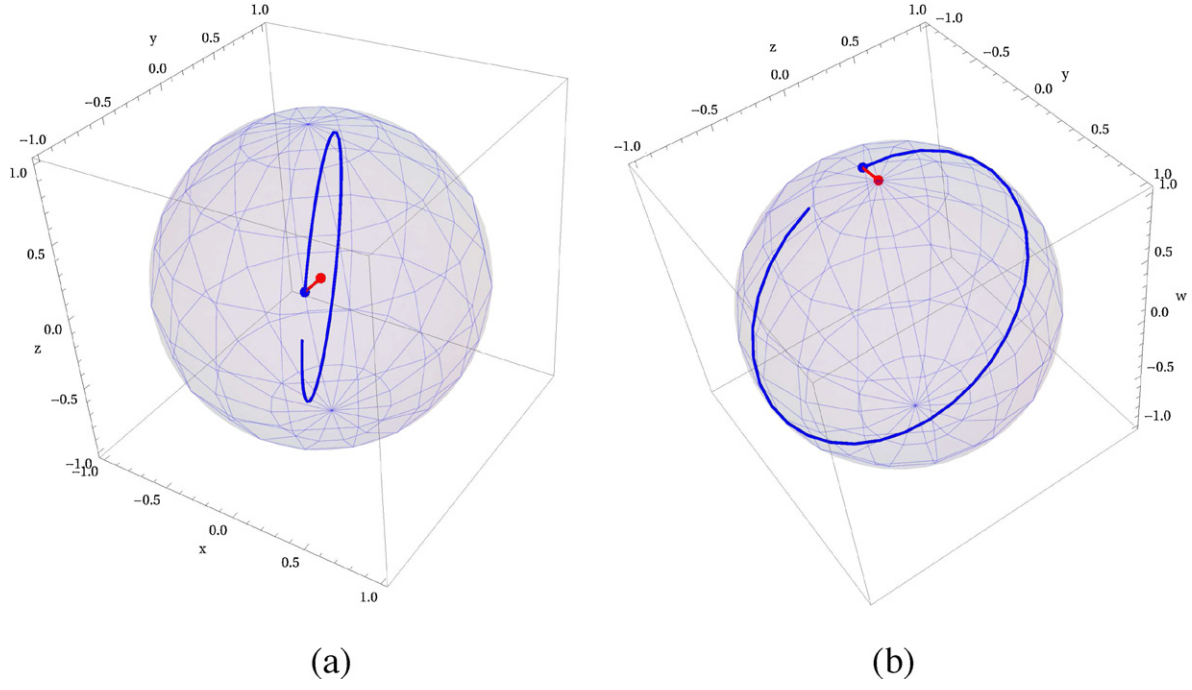


Fig. 8. The quaternion maps for a helix defined by the parametric equation $(r\cos(t), r\sin(t), pt)$. (a) The *xyz* quaternion map and (b) the *wyz* quaternion map of the continuous frames attached to the helix. The red line is the path from the identity frame (at the red dot) to the first actual helix frame. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

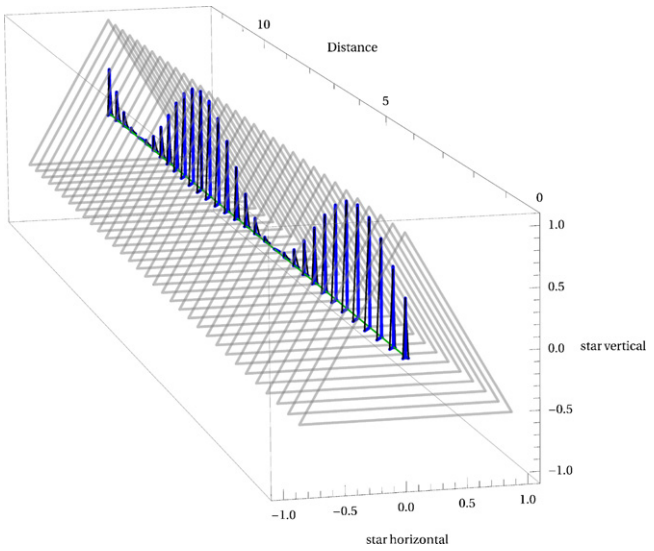


Fig. 9. Parallel *xyz*-squared star coordinates for the frames of a helix.

$$q_{\text{helix at 0}} = \left(\frac{1}{2}\sqrt{\frac{\sqrt{p^2+r^2}+r}{\sqrt{p^2+r^2}}}, \frac{p}{2\sqrt{r\sqrt{p^2+r^2}+p^2+r^2}}, \frac{-p}{2\sqrt{r\sqrt{p^2+r^2}+p^2+r^2}}, \frac{1}{2}\sqrt{\frac{\sqrt{p^2+r^2}+r}{\sqrt{p^2+r^2}}} \right). \quad (4)$$

Multiplying to the left by the quaternion $q_{z\,\text{rot}}(t) = (\cos(t/2), 0, 0, \sin(t/2))$ rotating about the z-axis, the full quaternion frame for the helix is then

$$q_{\text{helix}}(t) = q_{z\,\text{rot}}(t) \star q_{\text{helix at 0}}. \quad (5)$$

Plugging these values into the quadratic form in Appendix A one finds the matrix whose columns are the vectors in Eq. (3).

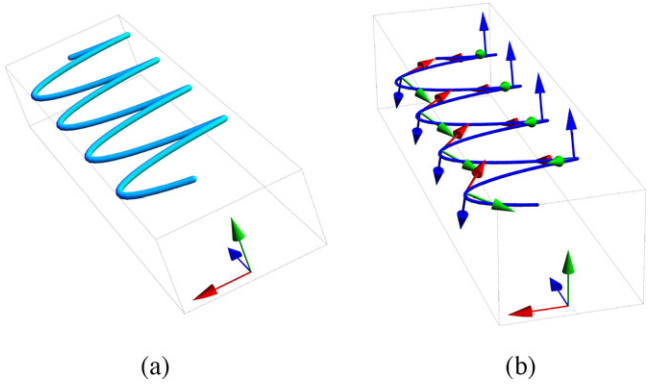


Fig. 10. (a) A beta sheet modeled by the parametric equation $(\cos(t), 0.1\sin(t), 0.5t)$. (b) A set of Frenet–Serret frames at roughly the expected places on the equation of the curve. Note the relation of the identity frame at foreground to the first actual sampled frame.

Fig. 8 shows the explicit helical quaternion maps in spherical geometric coordinates, in contrast to Fig. 9, which shows the star-squared parallel coordinate representation. The geometric forms in Fig. 8 are pure circles in the 4D geometry, and also in the *wyz* projection, but must be flattened ellipses in any other projection. The periodic circular path of the quaternion frame in Fig. 8 is reflected in the perfectly periodic pattern along the *z*-projection axis in the star-squared plot in Fig. 9.

4.2. Beta-sheet model: extreme quaternion frames

A crude mathematical approximation to a beta-sheet can be constructed by using a flattened helix such as Eq. (3) with $a \gg b$. Sampling the Frenet–Serret frames at $t = n\pi + \epsilon$ for a small random ϵ produces the alternating pairs of frame orientations characteristic of beta sheets. In real data, the beta sheet also twists systematically, which we could include in the model by a slow rotation in the *xy* plane. We show in Figs. 10 and 11 the beta-sheet analogs of the alpha helix model features shown in Figs. 7 and 8.

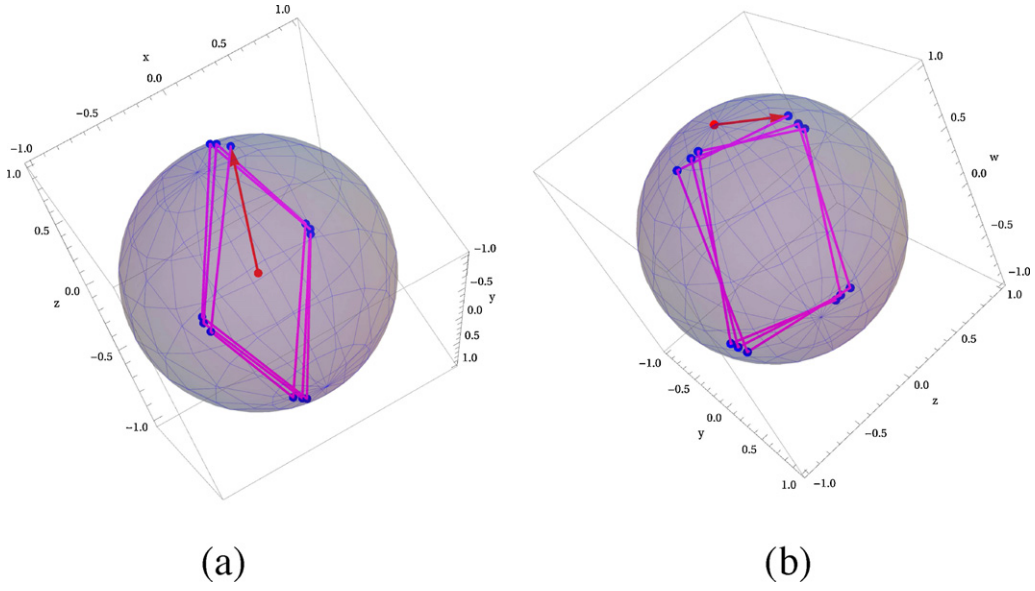


Fig. 11. A beta sheet modeled by the parametric equation $(\cos(t), 0.1\sin(t), 0.5t)$. (a) The corresponding *xyz* quaternion map of the continuous frames attached to the helix. The red arrow is the path from the identity frame (the red dot) to the first actual helix frame. (b) The *wyz* quaternion map (Note: the discontinuous nature of the beta-sheet frames is reflected in our choice of straight lines to connect neighboring frames here; in most cases, we will prefer to use smooth quaternion geodesics reflecting the shortest rotation path from one frame to the next). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

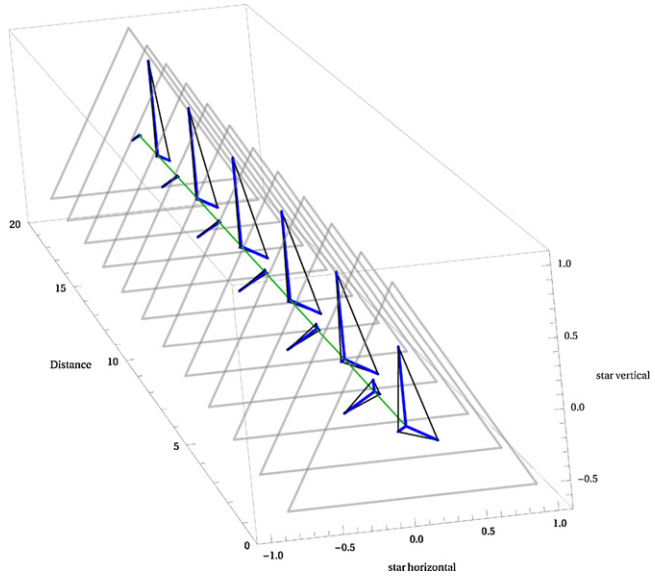


Fig. 12. Parallel *xyz*-squared star coordinates for the frames of an idealized beta-sheet model.

The parallel-coordinate-based star-squared map of the beta sheet model shown in Fig. 12 corresponds to Fig. 9 for the helix.

For idealized beta sheets, we observe clusters at intervals of roughly 90° in the quaternion plots, corresponding to the approximate 180° flips between neighboring residue orientations in a beta sheet. In practice, the real-world noisiness of the data will tend to interrupt the regular pattern of the mathematical model.

4.3. Quaternion frames from spline curves of PDB backbones

We next examine smooth frame sequences that can be associated directly with measured helical protein structures. Fig. 13(a) shows the structure of a helix-containing subsequence of a protein, the leucine zipper from the Protein Data Bank (PDB) file 1C94.pdb, whose dominant element is a single helical structure consisting of approximately seven loops. The idealized curve is defined by a smooth B-spline approximation to the path of the C_α atoms making up the backbone. This curve is continuously differentiable and is suitable for defining continuous moving frames along the curve such as the Frenet–Serret frames, samples of which are shown in Fig. 13(b). Fig. 13(c) is the quaternion *xyz* map of the Frenet–Serret frames for 1C94, showing the quaternion form of the sequence of orientations and their global relationship for the whole protein. Comparison with the pure mathematical helix in Figs. 7 and 8 clearly shows the close resemblance.

5. Quaternion frames from discrete PDB data

We turn our attention to quaternion descriptions of discrete 3D frames determined by exact atomic positions, rather than idealized curves. This will allow us to explore applications involving sequences of amino acid residue orientations.

There are many possible frame choices that can be assigned to components of a protein. We find it most natural to study those defined *within a residue*. Thus our prototypical frame will be the one anchored at the C_α carbon (C_α frame), shown in Fig. 14. Another useful but very distinct frame is the "P frame" (discussed below), which includes the direction of the peptide bond connecting a pair of residues, and thus utilizes atomic positions from both. The geometry of these frames is defined in detail in Appendix C.

There is one potential deficiency of the C_α frame, as pointed out in [18]: it is possible to fix both the absolute and relative orientation of two adjacent residues via their C_α frames, and *still* have a potentially significant ambiguity in the local geometric structure due to the so-called "bicycle motion" (see Fig. 15). The bond between the C_i atom and the N_{i+1} atom could possibly serve as the spinning crank joining two C_α-frame "bicycle pedals", though that action is severely limited by the rigidity of the peptide bond. This is of course true for *any* adjacent sets of three atoms in the protein backbone used to define independent frames.

In order to construct the protein geometry completely (up to whatever effect might arise from local distortions of the bond features), we would need at least one more intermediate frame such as the "P frame" relating two adjacent residues. The P-frame is shown schematically in Fig. 16 and defined in Appendix C. (Note that the Ramachandran angles, described in Appendix D, do not actually fully describe the transformation between adjacent frames.)

When we pass to sequences of discrete frames, remember that we must resolve the sign ambiguity between adjacent quaternion frame values by choosing the minimum quaternion distance to the preceding quaternion frame. For an ordered sequence of frames

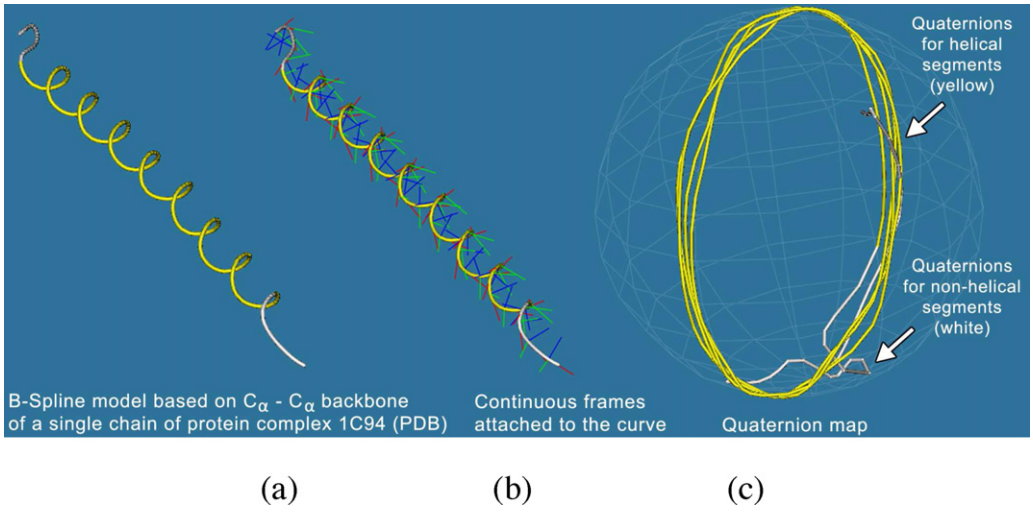


Fig. 13. A simple protein section, the leucine zipper in the 1C94 data set. (a) The standard B-spline curve derived from the underlying C_α backbone of the protein; note that this curve passes near, but not through, each C_α determined by this curve. (c) The *xyz*, quaternion map of the continuous Frenet–Serret frames. Compare these maps with those of the ideal helix in Figs. 7 and 8.

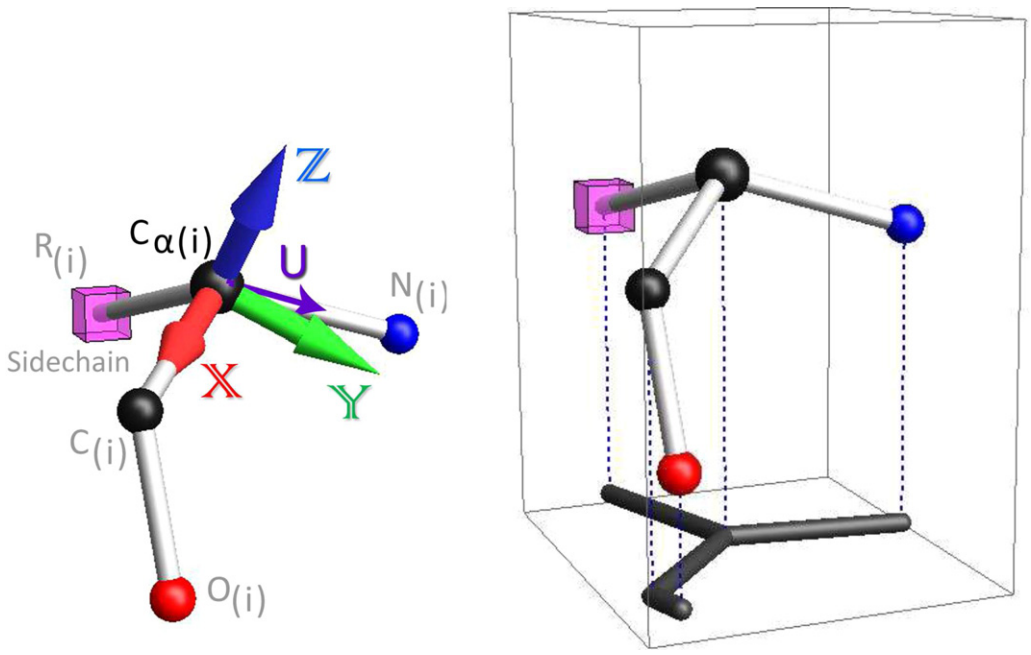


Fig. 14. Amino acid geometry showing the computation of our default discrete frame based on the direction from the C_α to the neighboring C and N atoms. The frame vectors **X** (red), **Y** (green), and **Z** (blue) are superimposed on the basic amino acid unit structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

such as those produced by protein residues, the resulting map is a sequence of points in $\mathbf{S}^3$ that can be connected by piecewise-continuous minimal-length quaternion curves [28] contained in the three-sphere, embedded in 4D Euclidean space, and projected according to the methods detailed above.

Our first example, the 1AIE structure, was introduced in Fig. 3 as a prototypical alpha helix. Applying the C_α frame map and the P frame map side by side, we find the results in Fig. 17, showing similar but not identical helical structures as ellipses in the spherical quaternion projection *wyz*.

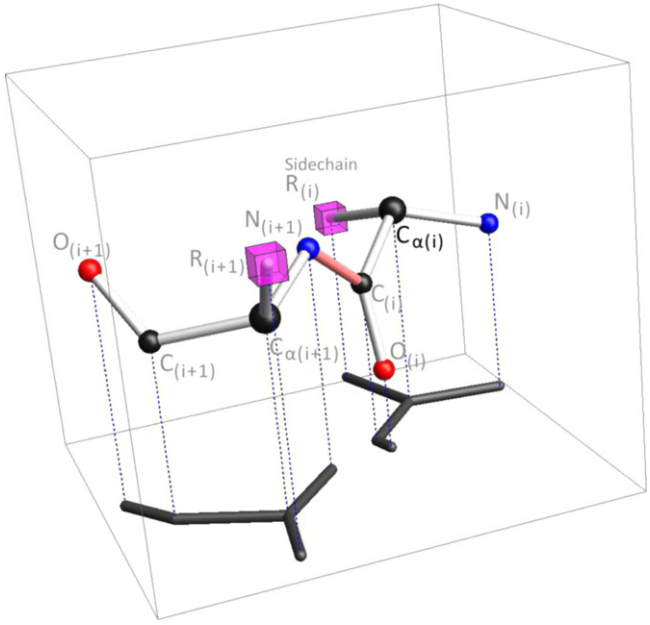


Fig. 15. Drop shadow representation of the geometry for two adjacent residues. The tinted C—N′ bond is central to the peptide bond, and embodies the "bicycle" ambiguity of the two neighboring C_α-frames. The P frame incorporates this peptide bond instead of being isolated in a single residue.

A more complicated configuration, the leucine zipper 1C94 doubled helix, is shown in Fig. 18, along with its C_α-frame quaternion maps. Each frame is represented by a single quaternion point in the map, and the ordered sequence of amino acids produces an ordered sequence of quaternion points. Any two points in this sequence, whether adjacent or not, can be connected pairwise by quaternion curves that correspond to the smallest rotation transforming one orientation frame to the other. The lengths of these minimal curves provide a precise measure of the similarity of the orientation frames. Amino acid residue frames that are close in quaternion space, whether nearby or distant in the ordered sequence, have similar global orientations.

Beta sheet example. We next examine the signature of beta sheets, which form widely spaced clusters of similar orientations in the quaternion maps, as shown in Fig. 19 for 2HC5, and later in Fig. 25. Our conventional coercion of neighboring quaternion frames to have positive inner products is not always effective for widely spaced frames such as those in beta sheets. In Fig. 20 we

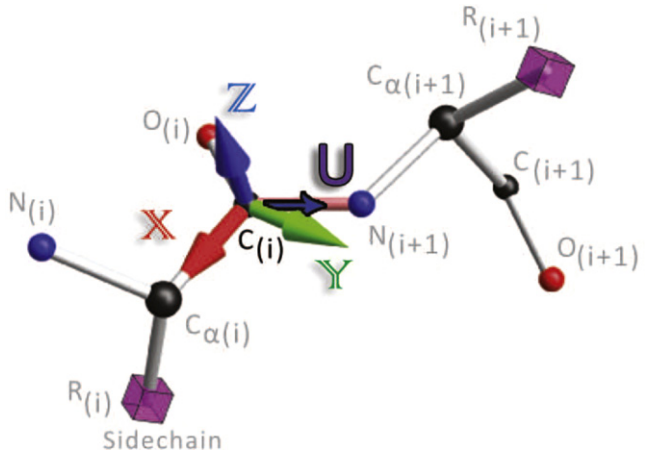


Fig. 16. The coordinates of the P-frame definition; the frame is centered on the C carbon, and extends to the nitrogen on the neighboring residue.

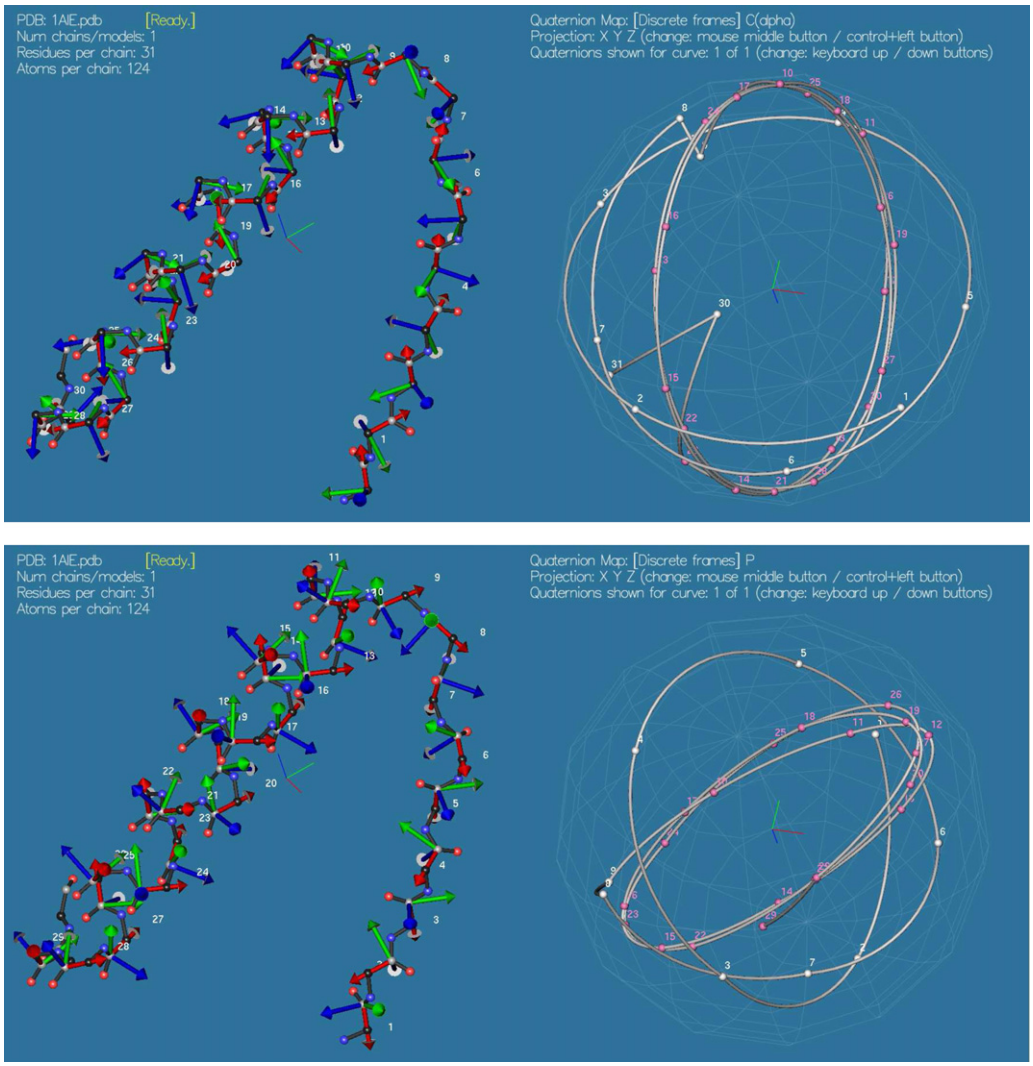


Fig. 17. Protein structure of 1AIE with its C_α frames and its P frames, and the corresponding quaternion frames joined as a sequence of spherical arcs.

show an alternative method, plotting *both* signs for frames suspected to form beta sheets, and clearly exhibiting the theoretically expected four-fold clustering.

Another type of example is shown in Fig. 21, which includes a B-Spline model of the protein 1A05, based on the C_α backbone. Orientation frames for residues that correspond to a beta sheet are drawn over the model shown in Fig. 21(a). A quaternion map in Fig. 21(b) shows quaternion points corresponding to the orientation frames. The quaternion map reveals that pairs of alternating residues have similar orientations. Some of the pairs of similar orientations are highlighted (A–D in the figure).

Observations. We can see by examining Figs. 3 and 17 that alpha-helices also produce clusters of similar orientations, and that every seventh amino acid frame falls close to its predecessor. For the particular case of 1AIE (see Fig. 3), the number of residues in the helix is small enough that we can single out seven distinct groups of two or three (marked by oval outlines in the figure) that are spaced seven apart in the sequence making up the helix. This is an example of the application of the quaternion map to highlight global orientation patterns that may be difficult to extract by other methods. In contrast, beta sheets will produce isolated clusters for short sequences, and more highly spread patterns for longer sequences. We can exploit the quaternion map in general to extract similarities in orientation patterns.

6. Example applications of discrete global quaternion frames

Applications of quaternion maps to the analysis of orientation frames fall into several categories:

- **Single or Composite Rigid Protein Frame Groupings:** The available data sets are dominated by explicit atomic locations for one single protein or a few closely associated proteins. The most useful information for such data sets is probably the set of discrete global frames based on a single residue, such as the C_α frame. Incorporating information from neighboring residues to form alternate sets of frames is possible, and can produce quaternion alternatives to the Ramachandran plot (see Appendix D). The backbone atoms can also be exploited to generate approximate polynomial curves representing protein structure; the analysis of such curves is exhaustively detailed elsewhere [22,23].

  We will focus on the residue-local C_α frame in our examples. Such discrete frames are particularly appropriate for identifying clusters of globally similar frames, which may be near one another physically or farther away but belonging to a regular geometric structure. Such clusters expose the natural relationships among groups of frames with diverse spatial relationships.
- **Patterns and Straightness:** Proteins arrange themselves into secondary geometric groups such as alpha helices and beta sheets.

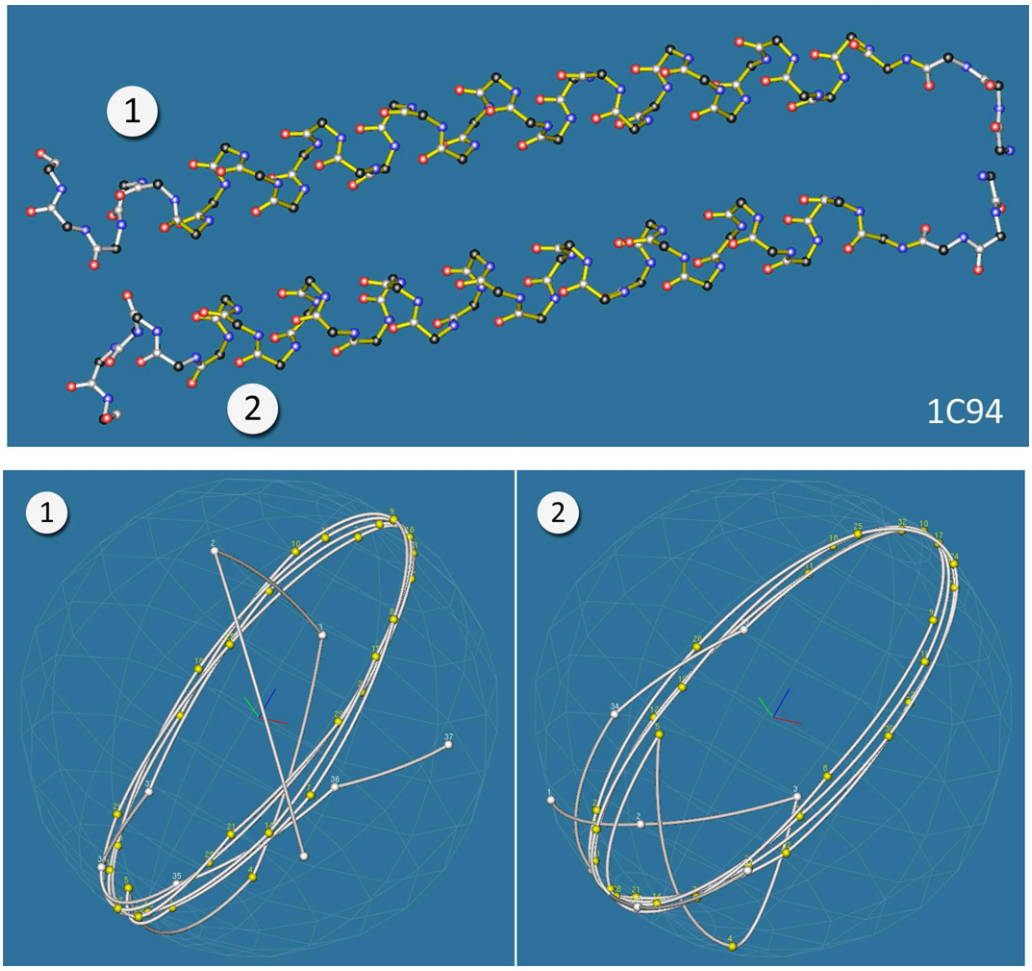


Fig. 18. (above) Geometry of the double-stranded protein structure 1C94 (the leucine zipper). Segments that are part of helices in the two strands are depicted in yellow. (below) Discrete quaternion maps for C_α frames of the two strands of 1C94. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Quaternions can be used for detailed analysis of the global orientations into patterns identified with secondary structures, and approaches have been found that use quaternion based "straightness measures" to effectively identify structural patterns [18].

- **Nonrigid Class Groupings:** The quaternion analysis does not depend on the rigidity assumptions underlying the X-ray crystallography data for atom locations in the PDB database. We can examine instead the nonrigid groupings of NMR data, which produce clusters of similar geometries for the same sequence of amino acid residues. These sheaf-like groupings of protein strands that present themselves in the NMR data provide an entirely different opportunity: here each individual amino acid appears multiple times, and quaternion measures provide essentially the only rigorous metric for quantifying the similarities of the orientations of the multiple instances of an amino acid in each of the strands. We employ both the spherical mean and the

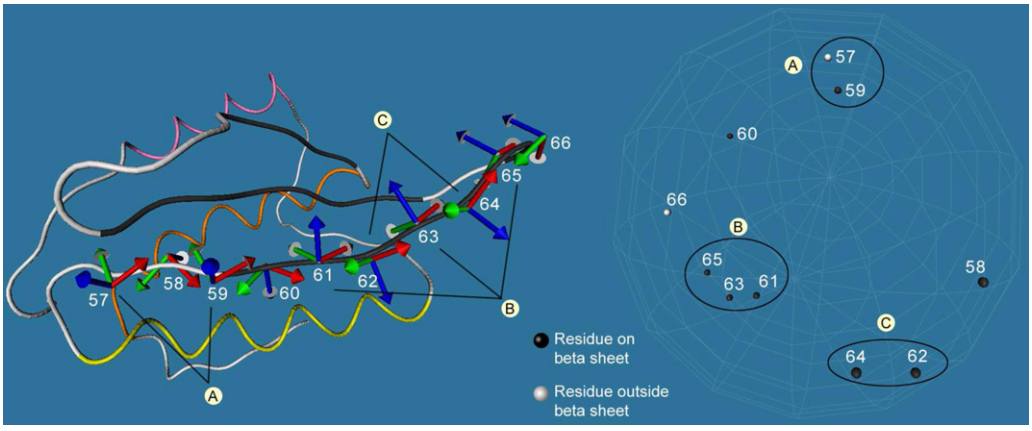


Fig. 19. Protein structure of 2HC5 and a quaternion map of its beta sheet structure. Neighboring frames are given matching quaternion signs in this map.

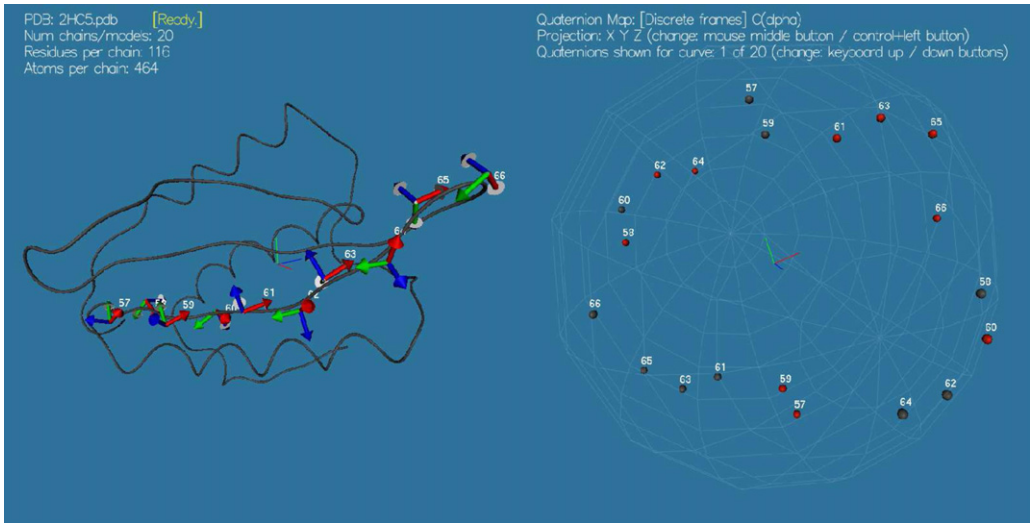


Fig. 20. Beta sheet structure with each quaternion frame displayed twice, with both possible overall signs. One can see that the beta sheet does not lie exactly in a plane, but twists slowly, causing the four expected 90° spaced groups to spread out across quaternion space.

standard deviation [29] to evaluate overall qualitative features of the cluster, and utilize outlier-excluding convex hulls for more robust descriptions of the low-rank statistics of these clusters. Examples are shown in Figs. 22–27.

- **Functional Activation Groupings:** Current research on enzyme functionality seeks to identify groups of active residues with side-chains that arrange themselves geometrically to facilitate biological functions. Given an hypothesis about, say, a triplet of side-chains we can compute quaternion representations not only for the basic C_α frame, but also for the orientations of relevant side-chains with respect to the C_α frame. Quaternion frames provide a relatively straightforward method for surveying proposed activation groupings for matching orientation patterns (see Figs. 28 and 29).

6.1. Quaternion frames of rigid proteins

Example: 1C94 (PDB), the Leucine Zipper. An elementary example is provided by a protein fragment known as the "Leucine Zipper," which consists of two α helices that align with one another in a tertiary structure (i.e., two or more associated proteins). The top of Fig. 18 shows the two strands and C_α—C_α backbones.

We compare the maps corresponding to the two different strands making up the protein complex of 1C94. The bottom of Fig. 18 shows quaternion maps for the two strands. While the maps for the two strands are expected to be similar because of the similarity of two strands, the maps can be used along with other metrics to uncover subtle differences, either within the same complex or across protein complexes.

6.2. Statistical quaternion groupings of NMR data

Example: NMR data for 1T50 (PDB): A Water-borne Pheromone from the Mollusk Aplysia Attractin. A more complex protein, the 1T50 pheromone, is depicted in Fig. 22(a). The structures in the figure correspond to a selection of twenty NMR data sets for the C_α—C_α backbone of the protein complex 1T50. Fig. 22(b) singles out one of these for reference. Fig. 22(c) shows the single model

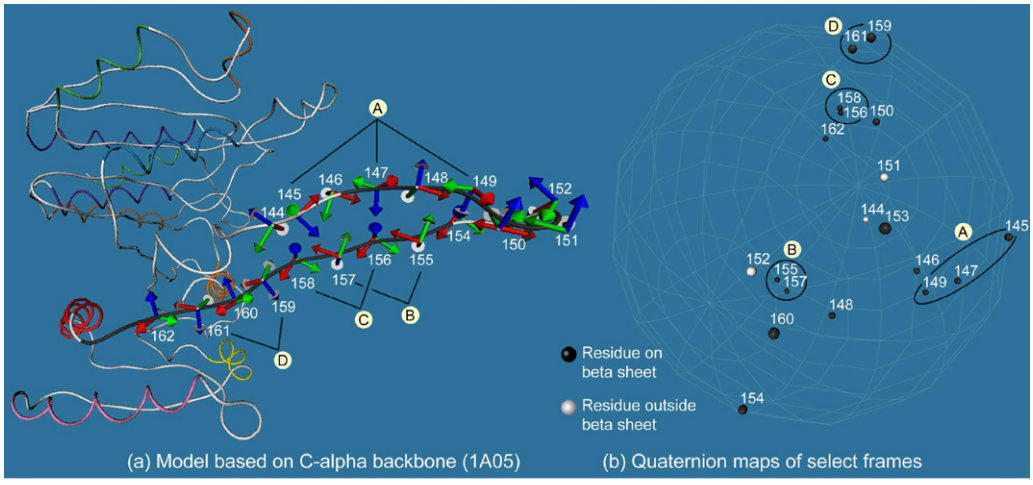


Fig. 21. (a) A model of protein 1A05 constructed using a B-Spline curve, which is based on backbone C_α atoms of the protein. The region with frames corresponds to the beta sheet structure. The frames are labeled by the sequence number of the amino acids they belong to. (b) Quaternion map of the select frames associated with discrete amino acids making up the protein 1AIE. Numbered points represent the quaternions corresponding to a single frame defined for each amino acid.

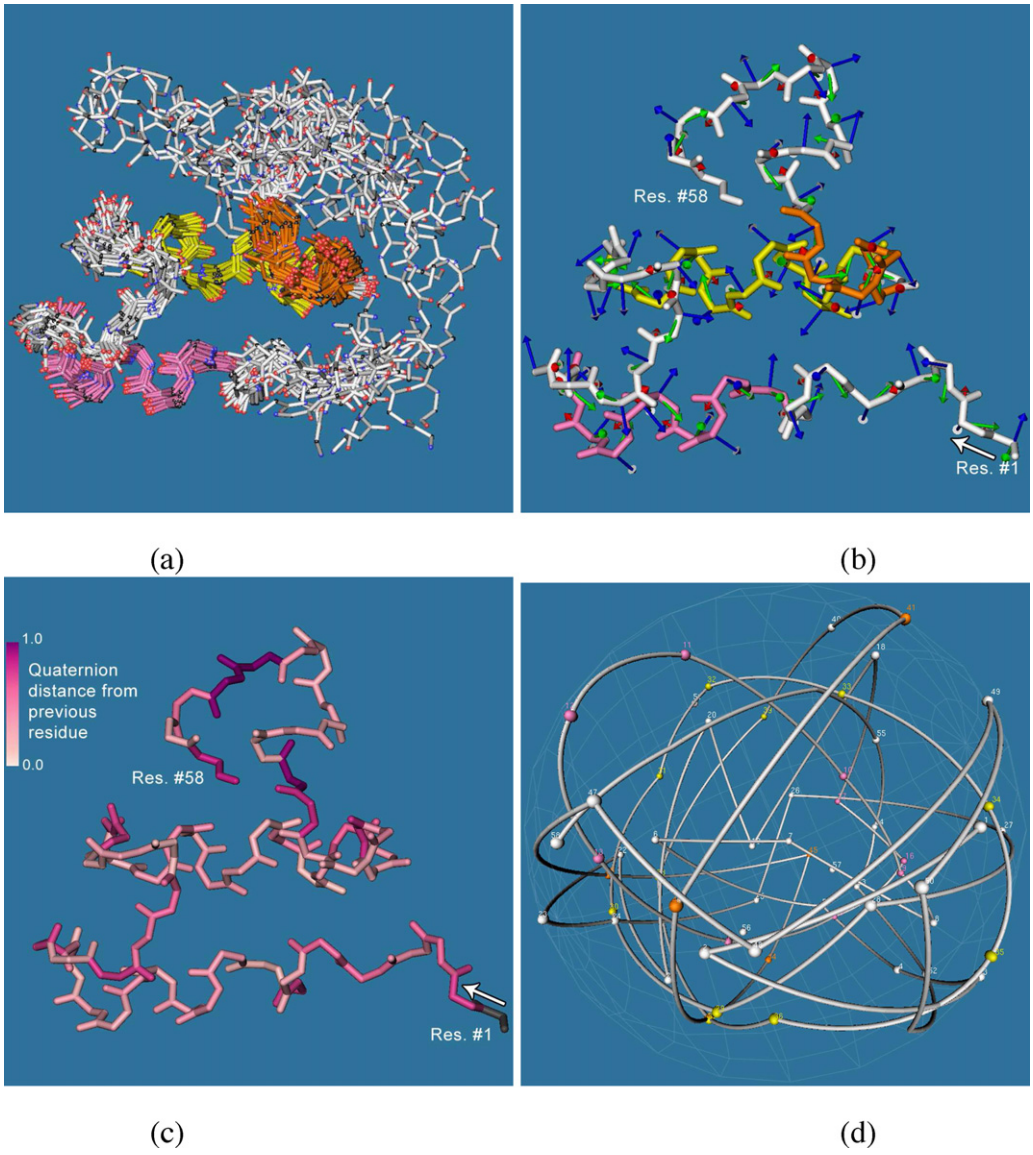


Fig. 22. Structure and quaternion map of the protein 1T50 (PDB), a water-borne pheromone. (a) Twenty models of the protein 1T50 determined using NMR structure, (b) backbone of a single model of the NMR dataset showing C_α frames, (c) the model shown in (b) color coded by quaternion distance of a residue relative to its preceding residue on the backbone, and (d) quaternion map of the frames shown in (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

color coded by quaternion distances between neighboring residues. The quaternion map for the reference segment (b) is displayed in Fig. 22(d).

Example: NMR data for 2HC5 (PDB). Twenty different geometric configurations of protein YvyC from Bacillus subtilis, 2HC5, are shown in Fig. 23. In this data set, the variations in the orientation of each amino acid can be clearly seen in the quaternion map, along with clusters of similar and dissimilar orientations. The spatial displacements in the data have only minimal correlation with the orientation displacements observed in the quaternion map; however, their cluster centers and statistical characteristics can be clearly identified, with very "floppy" arms of the protein generating large orientation variances, and relatively rigid branches keeping close to one another in quaternion space.

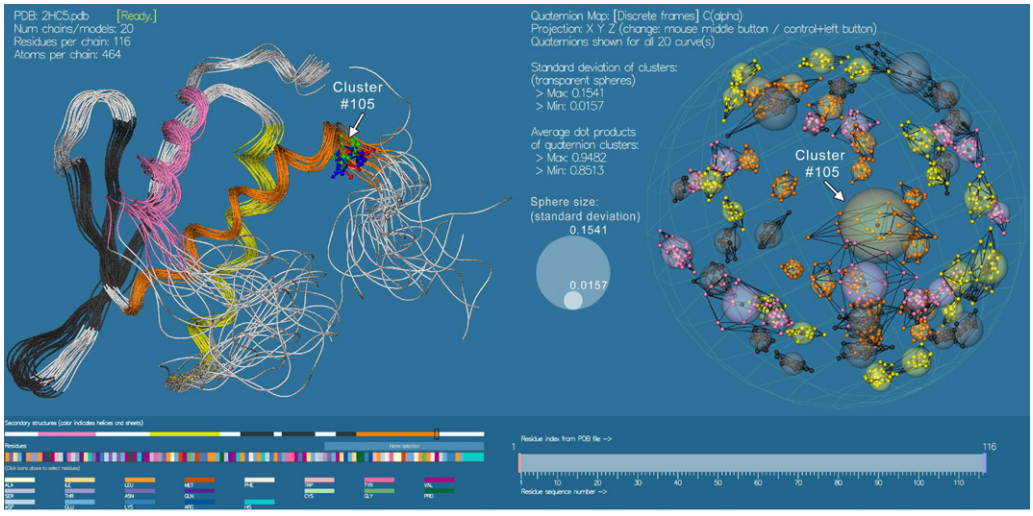
In Fig. 24, we interactively select a particular helical region of the protein to study the orientation distribution of its elements. Small quaternion regions correspond to fairly rigid configurations, and large regions have large quaternion-distance spreads around the spherical mean, indicating non-rigid behavior. Since the NMR data are selected on a relatively qualitative basis by the contributors, the precise meaning of some components of the differences are elusive, and it may well be possible to perform further quaternion-based analysis to further refine the apparent deviations in the data.

In Fig. 25, we show quaternion maps for alternative geometries of 1BVM and 2HC5. We note in particular the properties of the beta sheets that are components of these two structures.

Two particularly interesting examples of NMR analysis with quaternion visualization of the statistical distance distributions are given in Fig. 26, which shows a very "tight" distribution in the protein shape flexibility, and a contrasting situation in Fig. 27, which shows significant variation in the spatial structure, but retains relatively close distributions in the orientation frames (quaternion dot products within the 0.9 range).

6.3. Enzyme functional structures

Comparing HIS:TYR:ARG structures. It is known that catalytic residues can exhibit characteristic geometric structures [30]. Among the groups of structures that could have similar structure and behavior, the proteins 1CB8 and 1QAZ form an interesting pair



(a) (b)

Fig. 23. Quaternion maps for NMR data describing 10 different observed geometries for the protein yvyC from Bacillus subtilis, 2HC5. (a) The collection of alternative geometries. (b) Quaternion maps showing the *orientation space* geometry spreads for each individual amino acid.

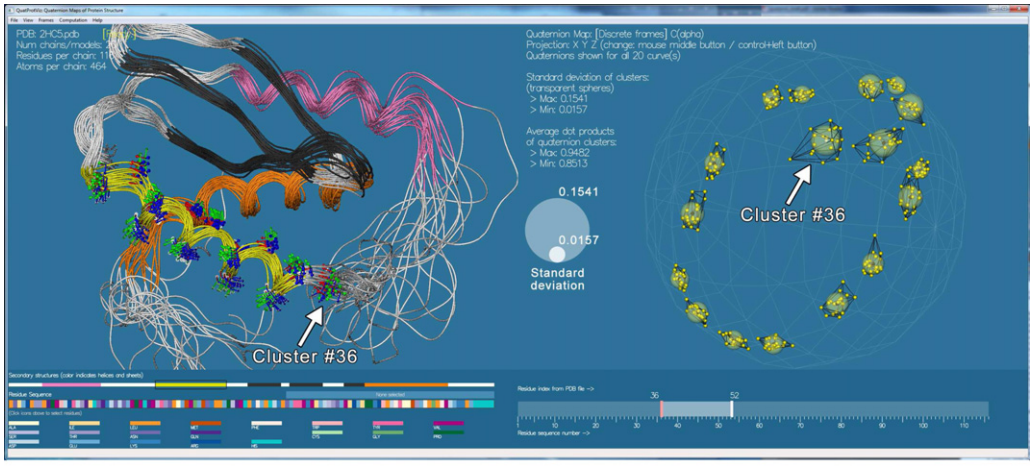
of examples, with similar physical locations of HIS, TYR, and ARG, with quaternion frames noted in the following table.

ID	Residue	(q_0, q_x, q_y, q_z)
1cb8	HIS 225	(0.861990, 0.491292, −0.124869, −0.003658)
	TYR 234	(0.146887, 0.799738, 0.297092, 0.500579)
	ARG 288	(−0.637504, 0.412136, −0.383576, 0.525929)
1qaz	HIS 192	(0.385861, −0.419153, −0.585444, 0.576781)
	ARG 239	(0.117073, 0.047674, 0.638640, 0.759052)
	TYR 246	(−0.565738, 0.038947, 0.598576, −0.565801)
1fmi	GLU 330	(0.443820, −0.519370, −0.290678, −0.669915)
	ASP 463	(−0.200536, 0.850733, 0.058078, 0.482355)
	GLU 599	(0.082620, −0.582741, −0.708285, 0.389768)
1rfn	HIS 57	(−0.154496, 0.404741, −0.857509, 0.277477)
	ASP 102	(−0.116981, 0.070199, −0.983358, −0.119979)
	SER 195	(−0.699076, −0.128476, 0.354901, −0.607316)

In Fig. 28, we show how these structures appear in 3D space. A different type of detail is shown in the quaternion plots: the objective in Fig. 29 is to successively align parts of the enzyme orientation in sequence to single out similarities and differences in these very distinct structures. Fig. 29(a) begins the process by showing the identified active sites listed above for the 4 enzymes 1cb8, 1qaz, 1fmi, and 1rfn, with the quaternion plots of their orientations all transformed to have the reference frame of the first residue as the origin (the identity frame). Technically, that means that all orientations have been multiplied by the inverse of the frame matrix of the first residue. Fig. 29(b) is the result of identifying the axis of rotation characterizing the rotation of the 2nd residue from the identity frame, and applying a global rotation on the entire enzyme to align that axis with the q_x axis. At this point, there is still one more degree of freedom, since the quaternion curve denoting the rotation from the 2nd residue's frame to the 3rd residue's frame can still be realigned.

7. Tools for exploring and comparing quaternion maps

Visual analysis of the quaternion maps can reveal interesting global information about the protein structures. However, there are several useful techniques at our disposal that can enhance the



(a) (b)

Fig. 24. Isolating a selected section of the protein YvyC from Bacillus subtilis, 2HC5. (a) The selected region. (b) Quaternion maps showing the *orientation space* geometry spreads for each individual amino acid in this region.

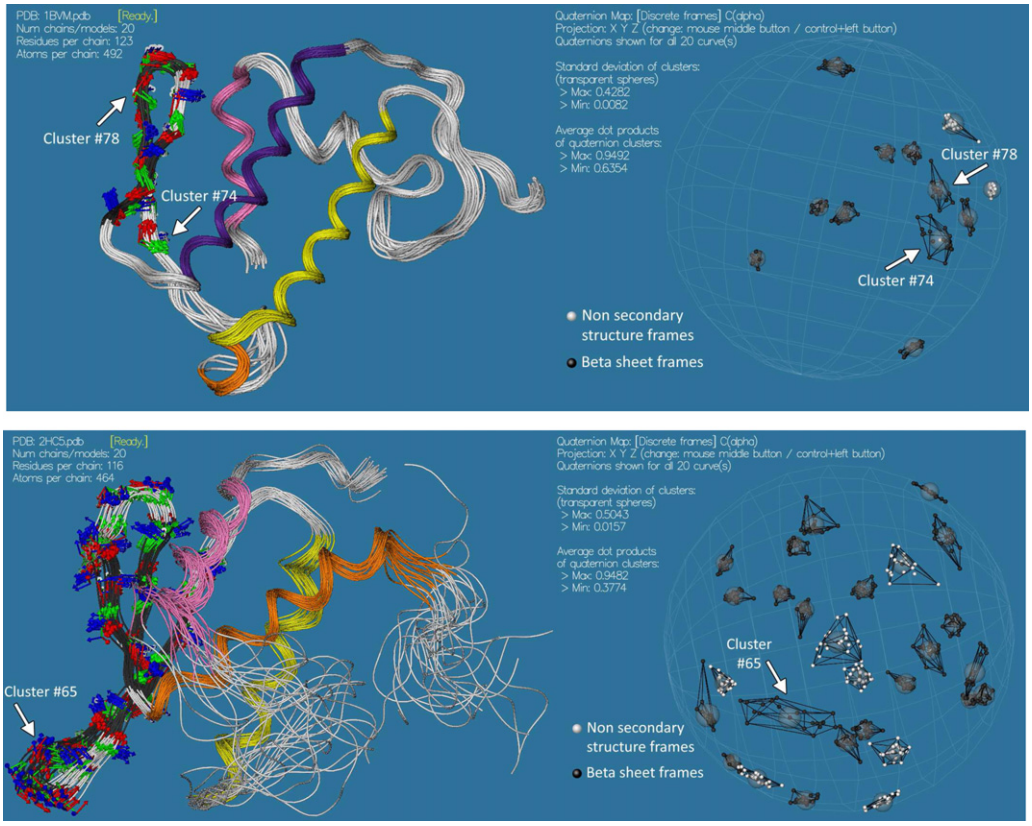
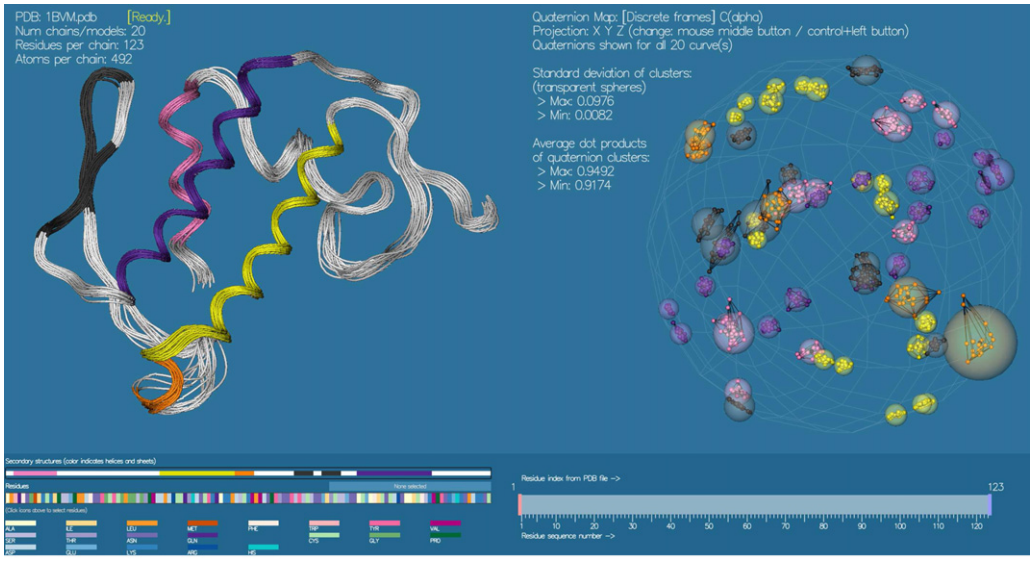


Fig. 25. NMR data for 1BVM and 2HC5, containing beta sheet examples.

visualizations or provide additional information, and can therefore aid in the exploration and comparison of the quaternion maps.

7.1. Aids for exploring quaternion maps

Our standard method relies on the fact that if we plot just the vector element $\mathbf{q}$ of the full unit quaternion $q=(q_0, \mathbf{q})$ obeying $q\cdot q=1$, then we have in principle a complete picture, since the fourth component $q_0=\pm\sqrt{1-\mathbf{q}\cdot\mathbf{q}}$ is redundant up to a sign. Curves, surfaces, and even volumes can be plotted in this way to show the global features of the quaternion orientation families and to represent available degrees of freedom. Several specialized techniques can further aid the visualizations:



(a) (b)

Fig. 26. Quaternion maps for NMR data describing 20 different geometries for the protein Bovine pancreatic phospholipase A2 derived from 1BVM.pdb. (a) The collection of alternative geometries. (b) Quaternion maps showing the *orientation space* geometry spreads for secondary structures of each of the predicted chains. This example has very close spatial similarity and quaternion frame similarity in the collection of alternative structures.

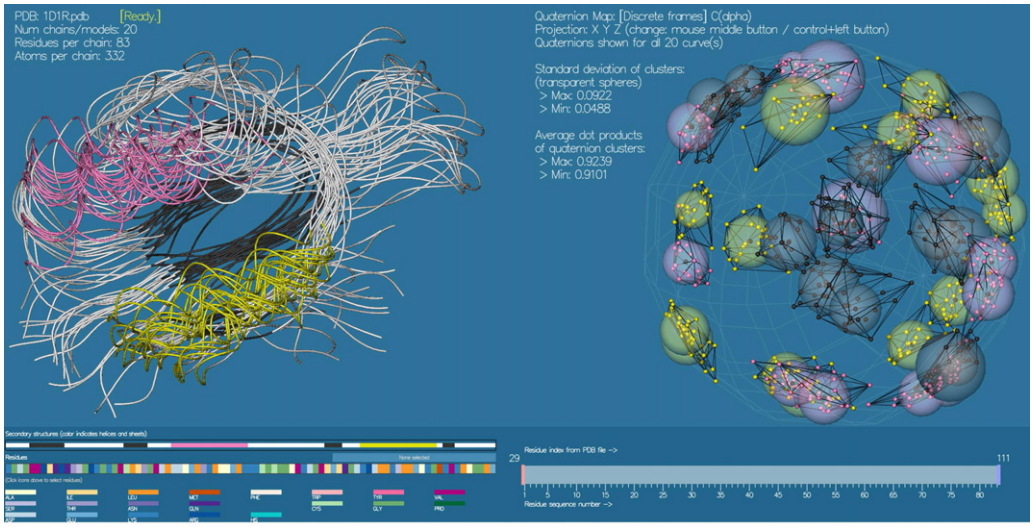


Fig. 27. Quaternion maps for NMR data describing 20 different geometries for the protein obtained from 1D1R.pdb; the protein is derived from genetic information in YciH gene of the E. Coli bacterium. (a) The collection of alternative geometries. (b) Quaternion maps showing the *orientation space* geometry spreads for secondary structures of each of the predicted chains. Note that even though the predicted structures in (a) are widely displaced in space, the error in orientations among corresponding residues is relatively low. It would be very hard to see this using any method except the quaternion plot.

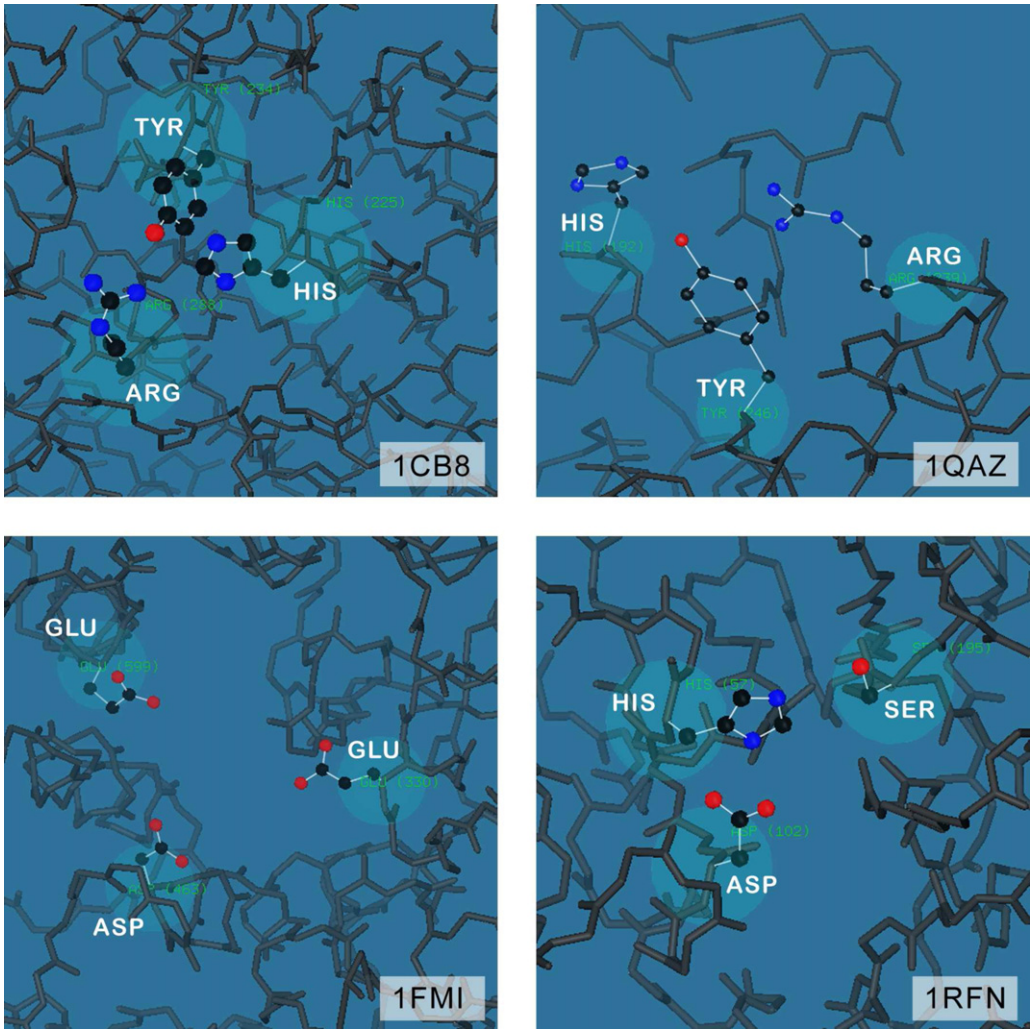


Fig. 28. The 3D locations of the active catalytic features of the proteins 1CB8, 1QAZ, 1FMI, and 1RFN.

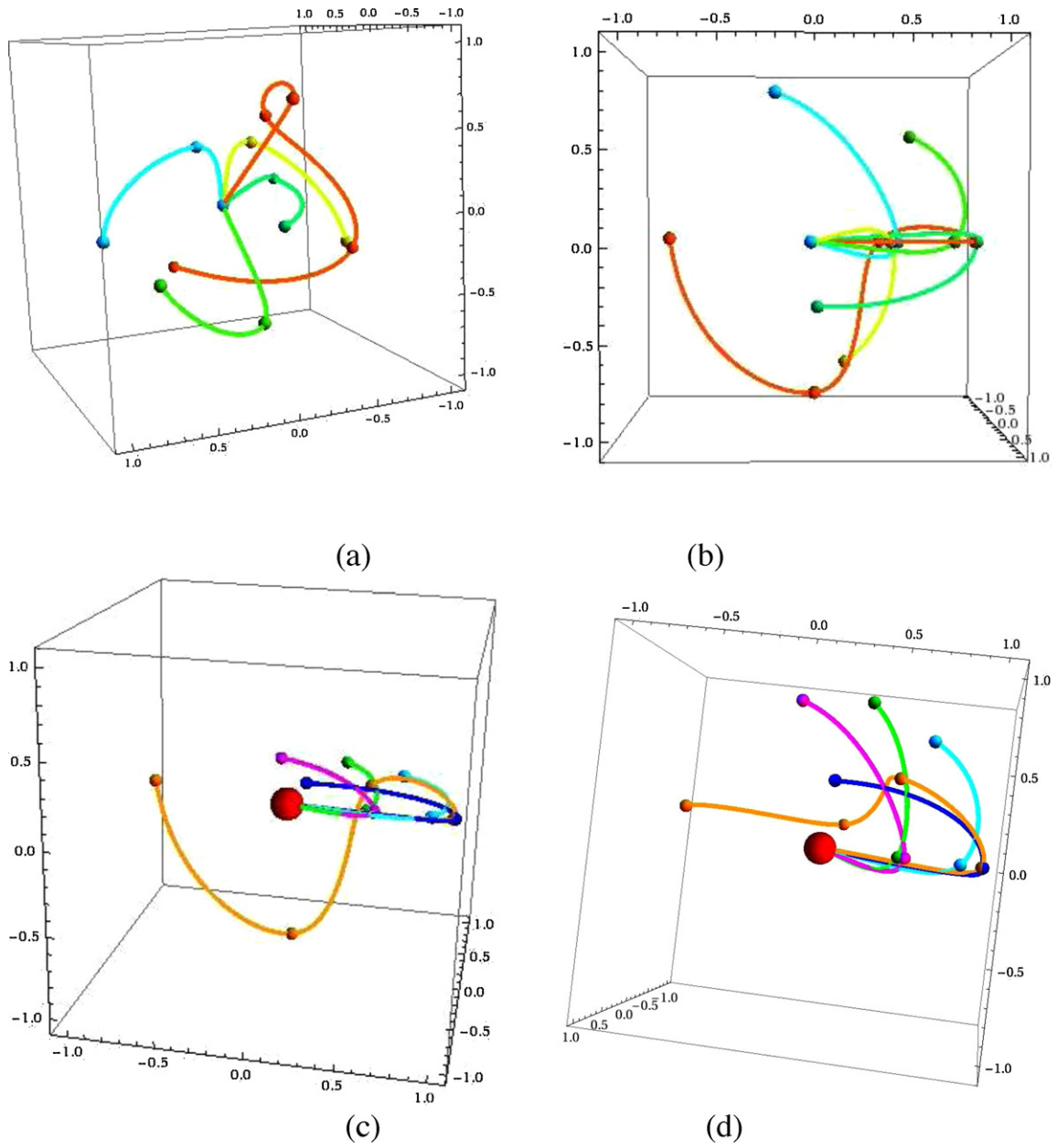


Fig. 29. (a) The quaternion maps of the listed active sites for 1IA6 (red), 1FMI (yellow), 1CB8 (green), 1QAZ (cyan), and 1RFN (blue), transformed to the same quaternion origin (first residue is the reference identity frame). (b) Result when the quaternion paths for all five enzymes from the first to the second residue are rotated to lie on the same axis. (c) Quaternion map that results when we perform the maximum possible alignment, with the frames of the 3rd residue rotated to lie in a common (q_x, q_y) plane. (d) Oblique view of the maximal quaternion space alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

- **Double frames.** The quaternion $q(F)$ corresponding to a given 3D frame F is ambiguous up to a sign. In principle, we should not be able to distinguish anything about the same family of frames if we assigned the sign at random. One approach to this feature is simply to place each frame twice in the plot, with both signs always present. For statistical clustering, and particularly for beta sheets, this can have some advantages; for applications dependent upon spatial sequencing of residues, this is less useful since any piecewise linear connections become very cluttered.
- **Force close frames.** One useful technique for studying sequences of quaternion frames is to assume that neighboring pairs are not wildly different in their orientations. Each pair of initially assigned neighboring quaternion frames can be characterized by the sign of their mutual dot product; the "force close frames" algorithm makes a smallest-rotation assumption, and changes the sign of the *next* quaternion frame if its dot product with its predecessor is initially negative. The result is a unique single sequence of quaternion frames, with no sign ambiguity or duplicated points, in which each neighboring pair of quaternion frames has a non-negative dot product, i.e., $q_{(i)} \cdot q_{(i+1)} \geq 0$.
- **Color coding.** While the value of q_0 is in principle superfluous, it can be useful to supply redundant information about its value, particularly in complicated long sequences, where q_0 may change sign several times. A simple way to do this is to color code the value of the unseen q_0 component at each plotted point of **q**.
- **Cycle through displayed quaternion components.** While our default *xyz* quaternion projection displays **q** and omits q_0, it may be useful to display q_0 as one of the three visible components and omit one of the q_k, e.g., the *wyz* projection. This is particularly useful for exposing certain types of circular or cyclic structures.

- **Grouping by skipping.** Many proteins have global orientation patterns that are not exactly sequential, but that may be exposed by sampling the protein at intervals. Thus if we group quaternion points corresponding to interval samplings, we can sometimes see the global orientation structure more clearly.

7.2. Metric for comparing quaternion maps

The quaternion maps for complex and large proteins can be dense, and discerning the relevant structure visually may become difficult. In such cases, we can use selection tools that rely on quaternion space distances to pull out various subsequences or similar regions.

Similar protein frames are characterized by quaternions that have larger mutual dot-products (cosine of the angle between 4-vectors closer to one) and so are closer in quaternion space. We can select locations on the protein whose orientations are similar to any given prototype point by thresholding the dot product.

Another measurement of the global properties of protein frames is the total turning along the helices. This is incrementally measured by the angle that takes one frame and rotates it into the other. This turning angle is given as usual by the quaternion-based measure computed from the dot product.

8. Conclusion

We have attacked the problem of defining global frames appropriate to the sequences of amino acids that make up proteins. Traditional methods for analyzing protein orientations such as the Ramachandran plot are useful for local relationships but have nothing to say about global orientation patterns or statistical distributions of absolute orientations. Since quaternion maps are precisely the right technology for revealing global orientation patterns and similarities, we have built a set of tools and methods that create quaternion maps of both discrete and continuous orientation sequences derived directly from the PDB file structure for any given protein with crystallographic of NMR based geometry. Quaternion maps thus provide a unique bridge between sequence and structure, and establish a tool that enables the asking of questions that have not previously been posed.

Future goals are to develop further tools to expose comparative global features, e.g., to see how different protein sequences may exhibit parallel geometric structures, and to analyze protein dynamics using quaternion tools.

Quaternion Tools. Demonstration tools for the investigation of quaternion frames are supported by the standard *jmol* environment, located at http://chemapps.stolaf.edu/jmol/. The *jmol* QUATERNION command can be used to generate and display quaternion maps for any PDB data.

Acknowledgements

This work was supported in part by National Science Foundation grant CCF-0204112. We are grateful to Tuli Mukhopadhyay for getting us started on the folklore of protein geometry, and particularly to Robert Hanson and his students for their creative contributions to the exploitation of quaternion protein technology. We benefited greatly from the generous suggestions and examples of enzyme structure hypotheses provided by Predrag Radivojac, Fuxiao Xin, and Yuzhen Ye.

Appendix A. Relating a quaternion to a frame

An orthogonal 3×3 matrix $\mathbf{F}$ can always be expressed as a rotation by an angle θ about a fixed direction $\hat{\mathbf{n}}$, $\mathbf{F} = \mathbf{R}(\theta, \hat{\mathbf{n}})$. The corresponding quaternion, up to an overall sign, is just

$$q = \left(\cos\left(\frac{\theta}{2}\right),\ \hat{\mathbf{n}} \sin\left(\frac{\theta}{2}\right) \right). \tag{A.1}$$

The matrix $\mathbf{R}$ can be expressed equivalently in terms of $(\theta, \hat{\mathbf{n}})$ or in terms of the quaternion components $q = (q_0, q_x, q_y, q_z)$: the $(\theta, \hat{\mathbf{n}})$ description is

$$\begin{bmatrix} c + (n_x)^2(1-c) & n_x n_y(1-c) - s n_z & n_z n_x(1-c) + s n_y \\ n_x n_y(1-c) + s n_z & c + (n_y)^2(1-c) & n_z n_y(1-c) - s n_x \\ n_x n_z(1-c) - s n_y & n_y n_z(1-c) + s n_x & c + (n_z)^2(1-c) \end{bmatrix}$$

where $c = \cos\theta$, $s = \sin\theta$, and $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$; the alternative quaternion expression is the quadratic form

$$\begin{bmatrix} q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2q_x q_y - 2q_0 q_z & 2q_x q_z + 2q_0 q_y \\ 2q_x q_y + 2q_0 q_z & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2q_y q_z - 2q_0 q_x \\ 2q_x q_z - 2q_0 q_y & 2q_y q_z + 2q_0 q_x & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix}.$$

These can be shown, with the help of Eq. (A.1), to be exactly the same thing. If the starting frame $\mathbf{F}$ is the identity matrix, the frame after transforming the coordinate system is given exactly by the *columns* of $\mathbf{R}(\theta, \hat{\mathbf{n}})$.

Since $\mathbf{R}(q)$ is quadratic in q, we see that $\mathbf{R}(q) = \mathbf{R}(-q)$, so q and $(-q)$ produce the same frame matrix. An equivalent statement is that the map is double valued, with $q(\theta)$ and $q(\theta + 2\pi)$ producing distinct opposite-sign quaternions but the same frame. When we multiply together two 3×3 rotation matrices ($\mathbf{R}_1$ composed with $\mathbf{R}_2$), the new frame $\mathbf{F} = \mathbf{R}_1 \cdot \mathbf{R}_2$ is exactly the same as the frame resulting from applying the quadratic map above to the quaternion

$$Q = q_1 \star q_2 = (w_1 w_2 - \mathbf{x}_1 \cdot \mathbf{x}_2,\ w_1 \mathbf{x}_2 + w_2 \mathbf{x}_1 + \mathbf{x}_1 \times \mathbf{x}_2)$$

where $\star$ is quaternion multiplication. The inverse q^{-1} of $q(\theta, \hat{\mathbf{n}})$ is just $q(-\theta, \hat{\mathbf{n}})$.

All that is required to convert a frame $\mathbf{F}$ to a quaternion is to find θ and $\hat{\mathbf{n}}$ of the corresponding rotation matrix $\mathbf{R}(\theta, \hat{\mathbf{n}})$ without encountering unacceptable numerical errors. Given $\mathbf{F}$, we can find θ and $\hat{\mathbf{n}}$ as follows:

$$\mathrm{tr}\mathbf{F} = 1 + 2\cos\theta$$

so

$$\cos^2\left(\frac{\theta}{2}\right) = \frac{1}{4}(\mathrm{tr}\mathbf{F} + 1).$$

Since

$$\mathbf{F} - \mathbf{F}^t = \begin{bmatrix} 0 & -2n_z \sin\theta & 2n_y \sin\theta \\ 2n_z \sin\theta & 0 & -2n_x \sin\theta \\ -2n_y \sin\theta & 2n_x \sin\theta & 0 \end{bmatrix}$$

we can in principle search for the largest value and then solve for $\hat{\mathbf{n}}$. When θ is near zero, one returns to the matrix $\mathbf{F}$ itself to find the largest off-diagonal term to use for a stable solution [23].

Appendix B. Quaternion distance

Distances between orientation frames are properly computed as the shortest paths lying within the quaternion three-sphere $\mathbf{S}^3$. This distance between two frames in isolation can be computed either from axis-angle rotation matrix methods or using quaternion methods. Placing these distances in a global context, however, requires quaternions. We can get a good estimate of the spherical distance from our 3D quaternion visualization methods, but there is

typically some spherical distortion in the projections that requires compensation using additional interactive tools. (This is similar to the problem of trying to make a distance-preserving projection of the Earth onto a flat piece of paper; distances between cities on a globe are perfectly correct, but we cannot take a satellite picture that allows accurate measurement of these distances with a simple ruler.) The distance of the frame defined by $\mathbf{R}(\theta, \hat{\mathbf{n}})$ from the identity frame may be understood in practice as the angle θ itself. We can express this distance in invariant form by noting that the 4D inner product (dot product) of the identity frame $q_{\mathrm{ID}} = (1, 0, 0, 0)$ with $q = (\cos(\theta/2), \hat{\mathbf{n}}\sin(\theta/2))$ is exactly $q \cdot q_{\mathrm{ID}} = \cos(\theta/2)$. The distance between any arbitrary pair of frames is thus seen to be

$$d_{12} = \theta_{12} = 2\cos^{-1}(q_1 \cdot q_2).$$

Since $q_1 \cdot q_2 = (q_1 \star q_2^{-1}) \cdot q_{\mathrm{ID}}$, some prefer to define this distance in terms of the first element of $q_1 \star q_2^{-1}$.

A practical representation of this minimal distance between two arbitrary quaternions q_1 and q_2 in $\mathbf{S}^3$ is the so-called SLERP [28]. This smooth minimal-length geodesic curve (which also projects to a smooth curve in any of our geometric views), is given by

$$S(q_1, q_2, t) = q_1 \frac{\sin((1-t)\phi)}{\sin\phi} + q_2 \frac{\sin(t\phi)}{\sin\phi} \tag{B.1}$$

where $\cos\phi = q_1 \cdot q_2$.

Appendix C. Protein frame geometry

In order to apply quaternion maps to a protein, we must identify full 3D coordinate frames that can be constructed uniquely from the atomic positions in a protein structure file, typically obtained from the Protein Data Bank [31], and derived from crystallographic or NMR data.

Each amino acid by itself defines a tetrahedral framework containing five atoms centered on the alpha-Carbon atom. Ignoring the hydrogen, we have the alpha-Carbon at the apex of a tetrahedron whose triangular base is formed by the C carbon of the carboxyl group, the side-chain R, and the nitrogen from the original amino group, as shown in Fig. 30. (Note that amino acids come in two isomers, D and L, and that most biological systems involve the L isomeric form shown in Fig. 30, where the left-handed CORN sequence points the thumb in the direction of the C_α.)

We can choose any three atoms to define a frame with one atom at the origin, and vectors from that central atom to the other

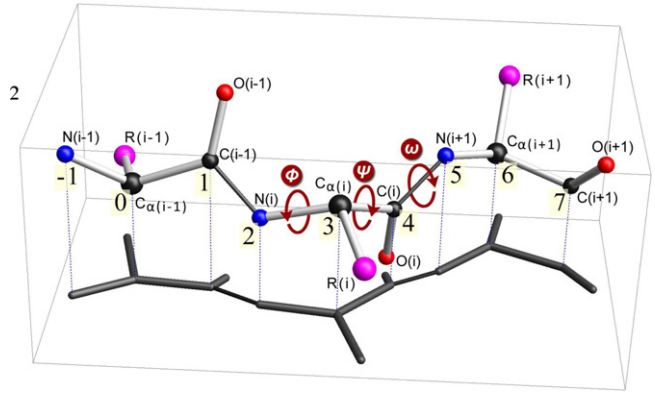


Fig. 30. Amino acid neighboring structure. Triples of atoms $(i-1)$, (i), and $(i+1)$ correspond to a single amino acid residue. The group of six atoms, C:N:C_α:C:N:C_α, starting at label "1" in the figure for the first C, defines the Ramachandran angles as "hinge" angles of the three groups of four atoms in the sequence of six. The planes of the peptide bonds connecting to adjacent amino acids define the ψ and ϕ dihedral angles. The angle ω describes the normally negligible torsion of the peptide bond, which is relatively rigid. The central tetrahedron has the alpha-Carbon at its top center, and we note that the orientation is the dominant L-form: the implicit hydrogen points upwards and the C—R—N triangle goes clockwise.

two serving to define the rest of the frame. Although we generally choose the C_α frame, with the triplet N:C_α:C forming the basis, we can choose any other local frame containing only atoms in a single residue, and it will be related to the C_α frame (or any other local frame) by a body-frame rotation that is universal for all amino acids (with the exception of a two-fold ambiguity for Glycine, for which the side chain (R) = H). We can also choose frames based on triples of atoms that cross residue boundaries; such frames form the basis of the Ramachandran angles.

Carbon alpha frame: For a local, single-residue frame, we therefore only need to define the C_α frame. Starting from the position vectors for each atom of the N:C_α:C triplet, we define the canonical C_α frame as follows:

$$\mathbf{X} = \frac{\mathbf{C}—\mathbf{C}_\alpha}{|\mathbf{C}—\mathbf{C}_\alpha|}$$
$$\mathbf{U} = \frac{\mathbf{N}—\mathbf{C}_\alpha}{|\mathbf{N}—\mathbf{C}_\alpha|}$$
$$\mathbf{Z} = \frac{\mathbf{X} \times \mathbf{U}}{|\mathbf{X} \times \mathbf{U}|}$$
$$\mathbf{Y} = \mathbf{Z} \times \mathbf{X}$$

as shown in Fig. 14.

Any such construction gives us a *frame* constructed from the fixed atomic vertices of an amino acid residue in a protein. The frame itself is representable as the 3×3 orthonormal matrix

$$\mathbf{F} = \left| \mathbf{X} \quad \mathbf{Y} \quad \mathbf{Z} \right|$$

and the corresponding quaternion $q(\mathbf{F})$ can be constructed (up to a sign) by the algorithm outlined above.

Peptide bond frame: The peptide bond frame, or "P frame," uses atomic positions from two neighboring residues sharing a peptide bond. Starting from the position vectors for each atom of the N:C_α:C triplet along with its following neighbor, N′:C'_α:C′, we define the canonical P frame as follows:

$$\mathbf{X} = \frac{\mathbf{C}_\alpha—\mathbf{C}}{|\mathbf{C}_\alpha—\mathbf{C}|}$$
$$\mathbf{U} = \frac{\mathbf{N}'—\mathbf{C}}{|\mathbf{N}'—\mathbf{C}|}$$
$$\mathbf{Z} = \frac{\mathbf{X} \times \mathbf{U}}{|\mathbf{X} \times \mathbf{U}|}$$
$$\mathbf{Y} = \mathbf{Z} \times \mathbf{X}.$$

The schematic image corresponding to the P-frame construction is shown in Fig. 16.

Side-chain frame: In addition to the main protein backbone coordinates, the PDB data files contain information on the positions of the atoms in the residue side-chains that can also be studied. Since the side-chain geometry and composition varies considerably, starting with the essentially structureless sidechain of glycine, which contains only a single hydrogen, one might need to customize the quaternion frame description on a case-by-case basis. Typically one would start with a framework such as the C:C_α:C_β triplet where C_β is the carbon atom on the side-chain group connected to C_α (if it exists). A prototype side-chain frame might then look something like

$$\mathbf{X} = \frac{\mathbf{C}_\beta—\mathbf{C}_\alpha}{|\mathbf{C}_\beta—\mathbf{C}_\alpha|}$$
$$\mathbf{U} = \frac{\mathbf{C}—\mathbf{C}_\alpha}{|\mathbf{C}—\mathbf{C}_\alpha|}$$
$$\mathbf{Z} = \frac{\mathbf{X} \times \mathbf{U}}{|\mathbf{X} \times \mathbf{U}|}$$
$$\mathbf{Y} = \mathbf{Z} \times \mathbf{X}.$$

Among other options, one could interchange C and C_β, or substitute N for C in the triad.

Appendix D. A quaternion context for traditional Ramachandran plots

In this appendix, we complete the overall picture that may be of interest to some readers, and describe in detail some relationships between the traditional 2D Ramachandran plot and our quaternion maps.

The standard triple of Ramachandran angles is determined by a sliding set of six atom positions as defined in Fig. 30. A convenient labeling, including the neighboring residues, is the following:

Atom:	N	C_α	C	N	C_α	C	N	C	C
ID number:	-1	0	1	2	3	4	5	6	7

standard frame

The Ramachandran starting position is the carbonyl carbon obtained by dropping the first two atomic positions (N and C_α) of the residue to the left of the residue that is our central focus, adjoining the $NC_\alpha C$ atoms of that residue, and appending the first two atoms (NC_α) of the residue to the right. We number this $CNC_\alpha CNC_\alpha$ sequence as 123456, with 234 being the atoms of the central residue, the one we have already used to define our standard quaternion frame parameters. The angle ϕ is associated with the 23 axis, ψ with the 34 axis, and ω with the 45 axis; however, we need to be careful about the signs, as described below.

In this group of six atoms, each set of three atomic positions from a PDB file defines a plane, and each pair of these triangles forms something that may be thought of as a bent hinge with the middle two atoms being the axis of the hinge (e.g., the vector $(\mathbf{3}-\mathbf{2})$ is the hinge of 1234). We may then label the normals to each of the triangles by the ordered triple of vertex indices (see Fig. 30), where we define the corresponding normal to be the result of the cross-product formed by the ordered vertices labeled as follows:

$$\hat{\mathbf{n}}(123) = \frac{(\mathbf{2}-\mathbf{1})\times(\mathbf{3}-\mathbf{2})}{|(\mathbf{2}-\mathbf{1})\times(\mathbf{3}-\mathbf{2})|}$$
$$\hat{\mathbf{n}}(234) = \frac{(\mathbf{3}-\mathbf{2})\times(\mathbf{4}-\mathbf{3})}{|(\mathbf{3}-\mathbf{2})\times(\mathbf{4}-\mathbf{3})|}$$
$$\hat{\mathbf{n}}(345) = \frac{(\mathbf{4}-\mathbf{3})\times(\mathbf{5}-\mathbf{4})}{|(\mathbf{4}-\mathbf{3})\times(\mathbf{5}-\mathbf{4})|}$$
$$\hat{\mathbf{n}}(456) = \frac{(\mathbf{5}-\mathbf{4})\times(\mathbf{6}-\mathbf{5})}{|(\mathbf{5}-\mathbf{4})\times(\mathbf{6}-\mathbf{5})|}.$$

Remember that $\hat{\mathbf{n}}(234)$ is the $\hat{\mathbf{z}}$-axis in Fig. 14. The cosine of each Ramachandran angle is given by the *inner* product of the pair of adjacent normals $\hat{\mathbf{n}}$, and the sign of the sine is given by the inner product of the hinge axis $\hat{\mathbf{A}} = \mathbf{A}/||\mathbf{A}||$ with the *cross* product of the two normals. Alternatively, writing $\mathbf{a}=\mathbf{2}-\mathbf{1}$, $\mathbf{b}=\mathbf{3}-\mathbf{2}$, and $\mathbf{c}=\mathbf{4}-\mathbf{3}$, and

$$x = (\mathbf{a}\times\mathbf{b})\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{b})(\mathbf{b}\cdot\mathbf{c}) - (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{b})$$
$$y = (\mathbf{a}\times\mathbf{b})\times(\mathbf{b}\times\mathbf{c})\cdot\frac{\mathbf{b}}{||\mathbf{b}||} = ||\mathbf{b}||(\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})),$$

we can determine the correctly signed cosine and sine from

$$\cos\phi = \frac{x}{\sqrt{x^2+y^2}}$$
$$\sin\phi = \frac{y}{\sqrt{x^2+y^2}}$$

where we cycle from 1234 through 2345 and 3456 to get ψ and ω, respectively.

While this basic geometry is well-known for the computation of the Ramachandran angles, we need the notation in order to proceed with the quaternion definitions that will allow us to gain some additional insights. First, we define the base coordinate system $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ as usual (from Fig. 14) for the vertices 234, with atoms $NC_\alpha C$. Then there exist three special rotations relative to that frame that we can write in axis-angle form as the 3D rotations

$$\mathbf{R}_1 = \mathbf{R}(\phi,\ \hat{\mathbf{A}}_1 = -\hat{\mathbf{A}}_{23})$$
$$\mathbf{R}_2 = \mathbf{R}(\psi,\ \hat{\mathbf{A}}_2 = +\hat{\mathbf{A}}_{34})$$
$$\mathbf{R}_3 = \mathbf{R}(\omega,\ \hat{\mathbf{A}}_3 = +\hat{\mathbf{A}}_{45}).$$

Here $\hat{\mathbf{A}}_{ij}$ is the normalized unit vector constructed from the atomic coordinates $\mathbf{j}-\mathbf{i}$, and the angles are the Ramachandran angles, the "hinge" angles of right-handed rotations leaving fixed the $\hat{\mathbf{A}}_{ij}$ axes. We need these because next we are going to define the corresponding quaternions whose positions in $\mathbf{S}^3$ represent the rotations that have to be applied to $\hat{\mathbf{n}}(234) = \hat{\mathbf{z}}$, the z-axis of our standard $NC_\alpha C$ frame, to change its direction to match the other three normals generated by the triangles in the 1234567 sequence.

$Q_1 = (\cos(\phi/2),\ \hat{\mathbf{A}}_1\sin(\phi/2))$	Rotates $\hat{\mathbf{z}}$ to align with the direction of $\hat{\mathbf{n}}(123)$, the inverse of the actual Ramachandran ϕ rotation.
$Q_2 = (\cos(\psi/2),\ \hat{\mathbf{A}}_2\sin(\psi/2))$	Rotates $\hat{\mathbf{z}}$ to align with the direction of $\hat{\mathbf{n}}(345)$, the Ramachandran ψ rotation.
$Q_3 = (\cos(\omega/2),\ \hat{\mathbf{A}}_3\sin(\omega/2))$	Rotates $\mathbf{R}_2\cdot\hat{\mathbf{z}}$ to align with the direction of $\hat{\mathbf{n}}(456)$, the normally ignored Ramachandran rotation that *precedes* the final rotation taking $\hat{\mathbf{z}}$ of the current $NC_\alpha C$ frame to $\hat{\mathbf{z}}'$, the z-axis of the *next* $NC_\alpha C$ frame in the protein.
$Q_4 = (\cos(\phi'/2),\ \hat{\mathbf{A}}_{56}[=-\hat{\mathbf{A}}_{1'}]\sin(\phi'/2))$	Rotates $\mathbf{R}_3\cdot\mathbf{R}_2\cdot\hat{\mathbf{z}}$ to align with the direction of $\hat{\mathbf{n}}(456) = \hat{\mathbf{z}}'$; this is the positive actual Ramachandran ϕ' rotation.

We can now choose an example protein representation, such as the PDB file for the mostly helical 1AIE with 31 residues, or the more complex 1A05 with 357 residues, and plot a variety of quantities for comparison.

- **Ramachandran angles.** We have ϕ and ψ, and so we can show the standard Ramachandran plots in Fig. 31, with clusters of points near $\phi \approx -60$ and $\psi \approx -40$ as is typical of the alpha helices contained in 1AIE and 1A05.
- **I: *xy* quaternion Cartesian sum map.** First we take the 3-vector parts of the quaternions Q_1 and Q_2 defined above and refer them to our standard C_α residue frame, so that the ϕ-rotation axis and the ψ-rotation axis lie in the same local reference frame, that is the local xy plane (by definition, the ψ-rotation axis is the x axis). The plot of these quantities in the 3D quaternion space as shown in Fig. 32, follows from simply adding the quaternion vectors, and this gives a quaternion-scaled 2D plot that is for all practical purposes indistinguishable from the Ramachandran plot. The most natural way to think of these 2D coordinates is as quaternion lengths arising from a single-axis rotation, they are also closely related (by replacing $\sin(\phi/2)$ with $\sin(\phi)$) to the axis-angle coordinates sometimes used in orientation analysis. The quaternions embed a rigorous metricity, while axis-angle coordinates are ad hoc.
- **II. *xy* quaternion product map.** The quaternion maps in Fig. 32 correspond essentially to the Ramachandran plot, and are constructed as a Cartesian sum of vectors that can be added in any order. This is not the way rotations actually act, and the rules of quaternion rotation representation are violated: while each single 3-vector Q_1 and Q_2 in Fig. 32 is a part of a unit length quaternion (remember that we can calculate the missing scalar part from the visible 3-vector), the Euclidean sum is not. However, we can correct that by performing a quaternion product, $Q_2 \star Q_1^{-1}$, and the result will be a quaternion that rotates the normal of the 123 triangle by the angle ϕ to the normal of the Frame 234, and then rotates that normal by ψ to align with the normal of

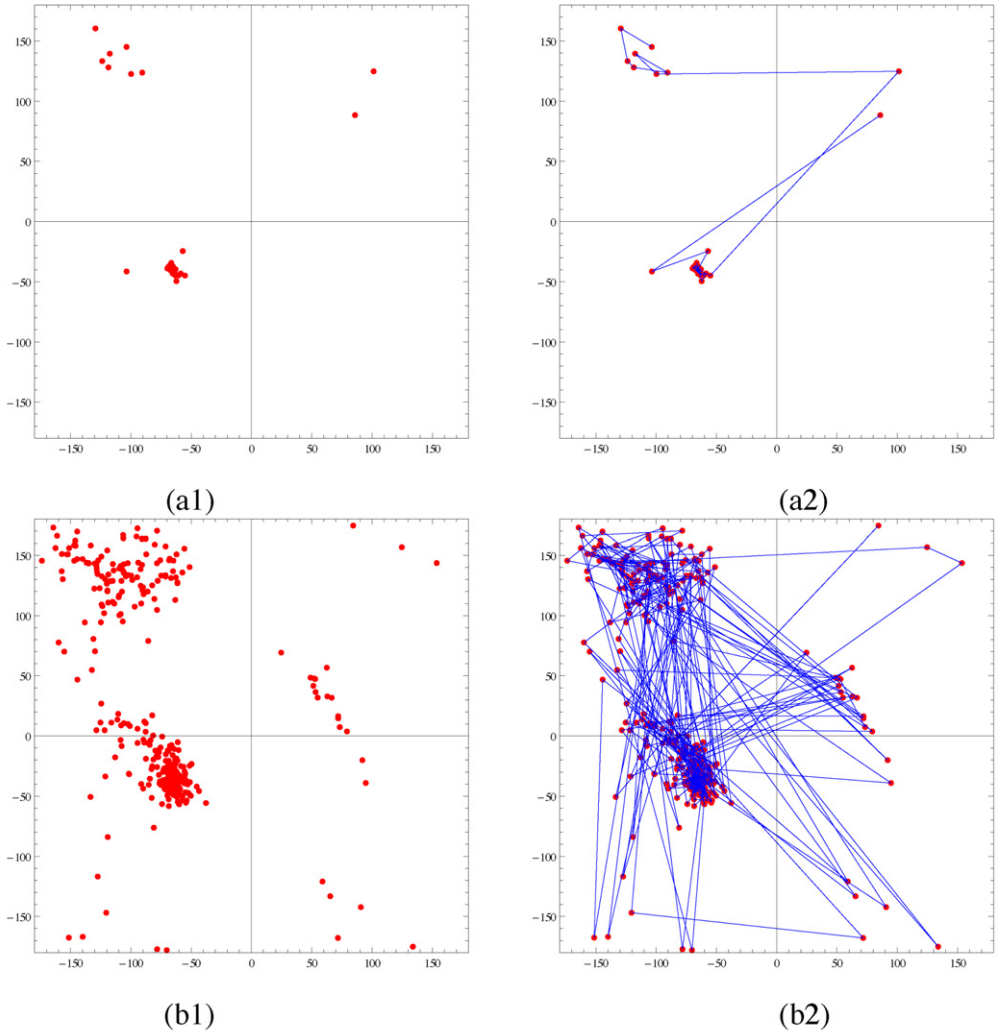


Fig. 31. (a1,a2) Standard Ramachandran plots of 31-residue 1AIE, with disconnected points and with adjacency-ordered line segments. (b1,b2) Same plots for the more extensive 357-residue 1A05 protein.

the 345 triangle; that is, the resulting quaternion represents the total rotation carried out when rotating by both Ramachandran angles to get approximately to the first "leaf" of the next ϕ frame.

Fig. 33 shows the results of this action on 1AIE and a 200 residue portion of 1A05, rotating the 123 normal until it aligns with the 345 normal. Reversing the order (distinct from using the inverse) results in a quaternion that differs by a sign in the z-component of the resulting quaternion.

- **III. Quaternion action of the three Ramachandran angles, and the missing twist.** The Ramachandran angles provide sufficient

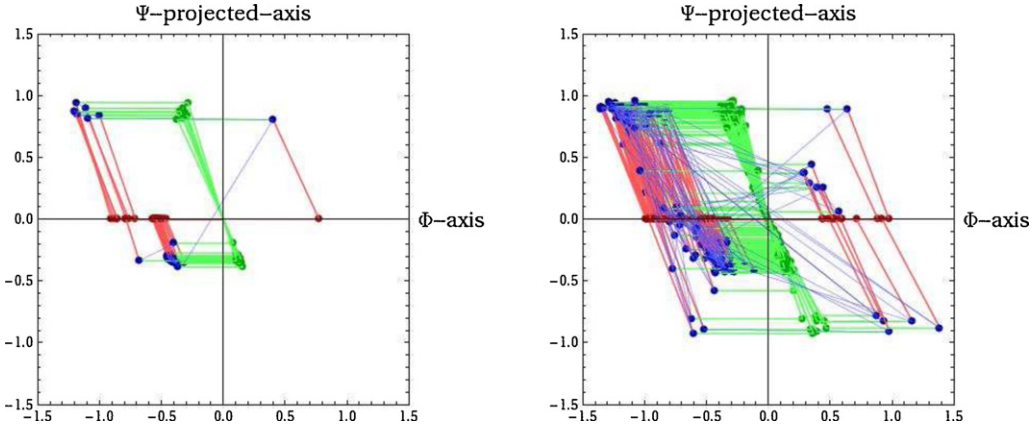


Fig. 32. Quaternion geometry corresponding to the standard Ramachandran plots for 1AIE (left) and 1A05 (right). Blue lines connect adjacent residues as in Fig. 31; red dots on the horizontal axis are the ϕ-related quaternion points, green dots are the ψ-related quaternion points relative to the quaternion coordinate frame. Blue dots are the Cartesian (Euclidean) sum of this pair of coordinates for each residue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

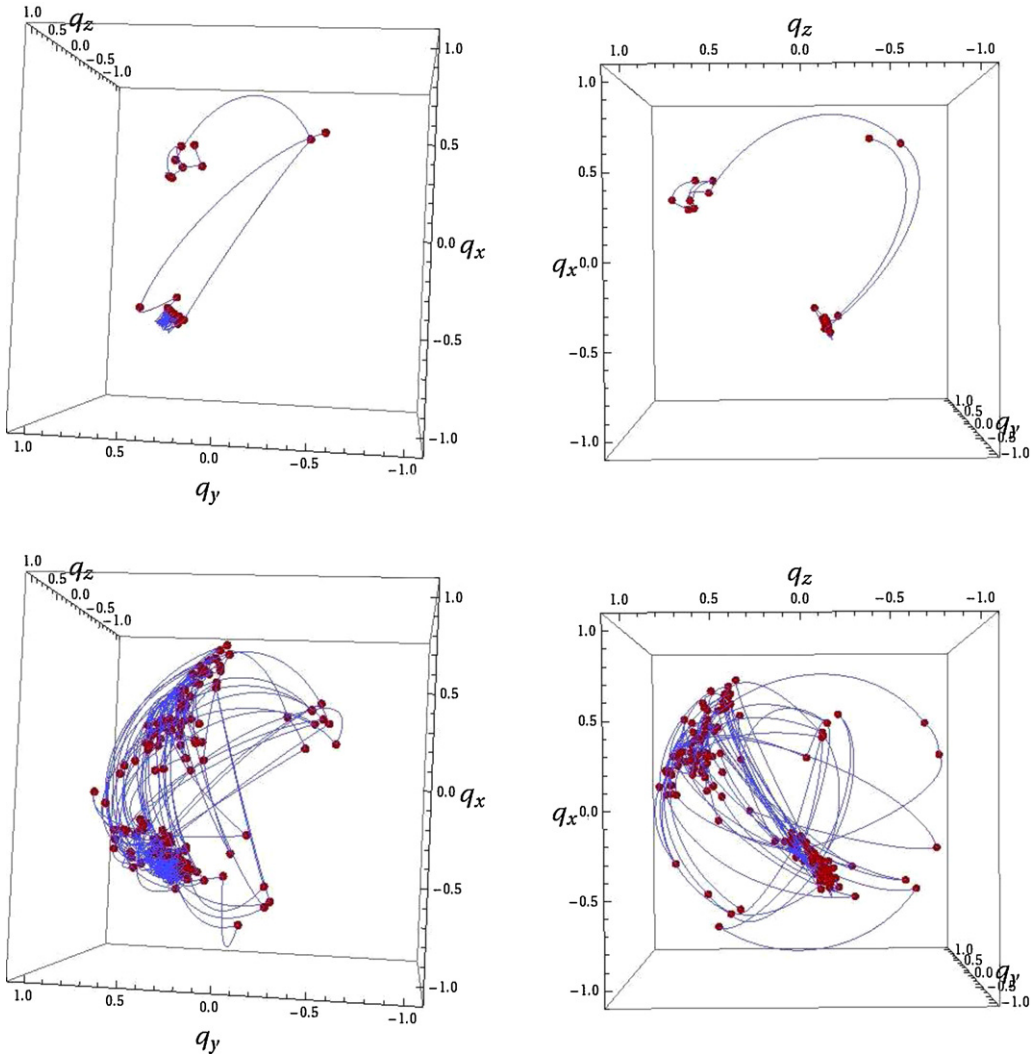


Fig. 33. Quaternion geometry with the action of the Ramachandran rotations represented as full quaternion products, with correct unit-length quaternion results, again for 1AIE (top, two viewpoints) and 1A05 (bottom, two viewpoints).

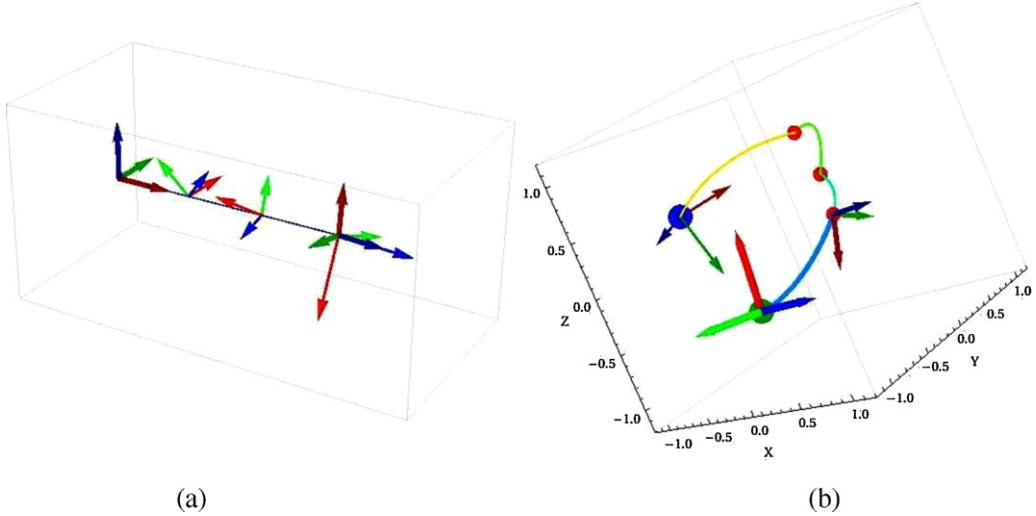


Fig. 34. Relationship between the rotations defined by the three Ramachandran angles taking one representative $NC_{\alpha}C$ frame to the next, and the quaternion value representing the transformation between the same two frames. The axial rotation performing the final frame alignment shown at the end of the path in the left image (a), and as a quaternion path in the right image (b); this orientation gap cannot be represented using the Ramachandran angles, which therefore lack crucial information.

information to define the transition from the plane of a given $NC_\alpha C$ frame to the next, provided we split them up so that the ψ and ω quaternions of the given frame are composed with the ϕ' frame of its successor. There are two steps needed to finally compare the Ramachandran data to the quaternion data in a fully quantitative fashion:

- We can find the value of the *new normal* for the next $NC_\alpha C$ frame, which we call $\hat{\mathbf{z}}'$, by applying the neighboring (split up) Ramachandran rotations in ordinary 3D space, and we can also express that complete rotation in quaternion form as follows:
  $\hat{\mathbf{z}}' = \mathbf{R}_{1'} \cdot \mathbf{R}_3 \cdot \mathbf{R}_2 \cdot \hat{\mathbf{z}}$
  $Q_{z\to z'} = Q_{1'} \star Q_3 \star Q_2$.
  However, all this tells us is the *orientation of the perpendicular* to the plane of the next $NC_\alpha C$ frame: it is powerless to tell us the *entire frame*. This is a deficiency of the Ramachandran approach.
- The final step necessary for complete understanding of the protein geometry, and one of our fundamental points in this treatment of protein orientation frames, is the addition of one final *spin* about the $\hat{\mathbf{z}}'$ axis! This is then the final relation between quaternion frames and Ramachandran angles: in Fig. 34(a), we show the location of a typical given $NC_\alpha C$ frame and use it as the identity reference frame, i.e., as a point at the origin of the *xyz* quaternion projection; then we plot the three quaternion arcs $Q_{1'} \star Q_3 \star Q_2$ in sequence taking that frame's normal $\hat{\mathbf{z}}$ to the next $\hat{\mathbf{z}}'$. But now we also plot the quaternion value of the *next* $NC_\alpha C$ frame, and see that it differs from the result of the Ramachandran transformation. The difference is simply a rotation by an angle σ about the $\hat{\mathbf{z}}'$ axis that can be computed in a number of ways, e.g.,
  $\mathbf{F}_{1'} = q(\sigma, \hat{\mathbf{z}}') \star Q_{z\to z'}$.
  where the quaternion frame $\mathbf{F}_{1'}$ or the $NC_\alpha C$ atoms having $\hat{\mathbf{z}}'$ as the normal to their plane, is computed in the coordinate system that has the original $NC_\alpha C$ frame $\mathbf{F}$ as the identity frame.

To conclude, in Fig. 34, we can lay out a plot of the global locations of all the quaternion frames for the entire protein in two equivalent forms: as the single quaternion arcs from $\mathbf{F}_i$ to $\mathbf{F}_{i+1}$, *or* as the *pair* of quaternion arcs consisting of the Ramachandran composite arc (from the total value of $Q_{z\to z'}$, composed with the $\hat{\mathbf{z}}'$-axis spin $q(\sigma, \hat{\mathbf{z}}')$). This last small "spin" arc is plotted in a thick line to emphasize the distinction between the global frame orientation and the information available from the Ramachandran angles.

References

[1] C.F. Karney, Quaternions in molecular modeling, Journal of Molecular Graphics and Modelling 25 (5) (2007) 595–604, http://dx.doi.org/10.1016/j.jmgm.2006.04.002.
[2] S.K. Kearsley, On the orthogonal transformation used for structural comparisons, Acta Crystallographica Section A 45 (2) (1989) 208–210.
[3] E.A. Coutsias, C. Seok, K.A. Dill, Using quaternions to calculate RMSD, Journal of Computational Chemistry 25 (2004) 1849–1857.
[4] A.L. Mackay, Quaternion transformation of molecular orientation, Acta Crystallographica Section A 40 (1984) 165–166.
[5] B. Horn, Closed-form solution of absolute orientation using unit quaternions, Journal of the Optical Society 4 (4) (1987) 629–642.
[6] Y. Magarshak, Quaternion representation of RNA sequences and tertiary structures, Biosystems 30 (1-3) (1993) 21–29, http://dx.doi.org/10.1016/0303-2647(93)90059-L, http://www.sciencedirect.com/science/article/pii/030326479390059L
[7] K. Albrecht, J. Hart, S. Alex, A.K. Dunker, Quaternion contact ribbons: a new tool for visualizing intra- and intermolecular interactions in proteins, Pacific Symposium on Biocomputing 94 (1996) 41–52.
[8] A.K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, Protein disorder and the evolution of molecular recognition: theory, predictions, and observations, Pacific Symposium on Biocomputing 3 (1998) 473–484.
[9] E.A. Coutsias, C. Seok, K.A. Dill, Using quaternions to calculate RMSD, Journal of Computational Chemistry 25 (2004) 1849–1857.
[10] D.J. Siminovitch, Rotations in NMR: part I. Euler–Rodrigues parameters and quaternions, Concepts in Magnetic Resonance 9 (1997) 149–171.
[11] D.J. Siminovitch, Rotations in NMR: part II. Applications of Euler–Rodrigues parameters, Concepts in Magnetic Resonance 9 (1997) 211–225.
[12] J.R. Quine, Helix parameters and protein structure using quaternions, Journal of Molecular Structure 460 (1999) 53–66.
[13] R. Srinivasan, V. Geetha, J. Seetharaman, S. Mohan, A unique or essentially unique single parametric characterisation of biopolymetric structures, Journal of Biomolecular Structure and Dynamics 11 (1993) 583–596.
[14] G.R. Kneller, P. Calligari, Efficient characterization of protein secondary structure in terms of screw motions, Acta Crystallographica. Section D, Biological Crystallography 62 (2006) 302–311.
[15] C. Branden, J. Tooze, Introduction to Protein Structure, 2nd ed., New York/London, Garland Publishing, 1999.
[16] G.N. Ramachandran, C.S.V. Ramakrishnan, Stereochemistry of polypeptide chain configurations, Journal of Molecular Biology 7 (1963) 95–99.
[17] V. Bojovic, I. Sovic, A. Bacic, B. Lucic, K. Skala, A novel tool/method for visualization of orientations of side chains relative to the protein's main chain, in: MIPRO'11, 2011, pp. 242–245.
[18] R. Hanson, D. Kohler, S. Braun, Quaternion-based definition of protein secondary structure straightness and its relationship to Ramachandran angles, Proteins: Structure, Function, and Bioinformatics 79 (7) (2011) 2172–2180, http://dx.doi.org/10.1002/prot.23037.
[19] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton, Stereochemical quality of protein structure coordinates, Proteins 12 (1992) 345–364.
[20] A.J. Hanson, H. Ma, Visualizing flow with quaternion frames, in: VIS'94: Proceedings of the Conference on Visualization '94, IEEE Computer Society Press, Los Alamitos, CA, USA, 1994, pp. 108–115.
[21] A.J. Hanson, H. Ma, Quaternion frame approach to streamline visualization, IEEE Transactions on Visualization and Computer Graphics 1 (2) (1995) 164–174, http://dx.doi.org/10.1109/2945.468403.
[22] A.J. Hanson, Constrained optimal framings of curves and surfaces using quaternion gauss maps, in: Proceedings of Visualization '98, IEEE Computer Society Press, 1998, pp. 375–382.
[23] A.J. Hanson, Visualizing Quaternions, Morgan Kaufmann, 2006.
[24] A. Inselberg, Parallel Coordinates: Visual Multidimensional Geometry and its Applications, Springer, 2009.
[25] J.M. Chambers, W.S. Cleveland, B. Kleiner, P.A. Tukey, Graphical Methods for Data Analysis, Wadsworth, Belmont, CA, 1983.
[26] E. Fanea, S. Carpendale, T. Isenberg, An interactive 3d integration of parallel coordinates and star glyphs, in: J. Stasko, M. Ward (Eds.), Proceedings of the IEEE Symposium on Information Visualization, InfoVis 2005, October 23–25, 2005, Minneapolis, Minnesota, USA, IEEE Computer Society, Los Alamitos, CA, 2005, pp. 149–156, http://dx.doi.org/10.1109/INFOVIS.2005.5.
[27] X.Q. Yan, C.W. Fu, A.J. Hanson, Multitouching the fourth dimension, Computer 45 (9) (2012) 80–88.
[28] K. Shoemake, Animating rotation with quaternion curves, in: SIGGRAPH'85: Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques, ACM Press, New York, NY, USA, 1985, pp. 245–254, http://dx.doi.org/10.1145/325334.325242.
[29] S.R. Buss, J. Fillmore, Spherical averages and applications to spherical splines and interpolation, ACM Transactions on Graphics 20 (2001) 95–126.
[30] F. Xin, S. Myers, Y.F. Li, D.N. Cooper, S.D. Mooney, P. Radivojac, Structure-based kernels for the prediction of catalytic residues and their involvement in human inherited disease, Bioinformatics/Computer Applications in The Biosciences 26 (2010) 1975–1982, http://dx.doi.org/10.1093/bioinformatics/btq319.
[31] The RCSB Protein data bank: site functionality and bioinformatics use cases, 2011, http://dx.doi.org/10.1038/pid.2011.1