Anthony K. Yan
Christopher J. Langmead
Dartmouth Computer Science Department
Hanover, NH 03755, USA

Bruce Randall Donald
6211 Sudikoff Laboratory
Dartmouth Computer Science Department
Hanover, NH 03755, USA
(also at Dartmouth Chemistry Department and
Dartmouth Department of Biological Sciences
Hanover, NH 03755, USA)
brd@cs.dartmouth.edu

A Probability-Based Similarity Measure for Saupe Alignment Tensors with Applications to Residual Dipolar Couplings in NMR Structural Biology

Abstract

High-throughput nuclear magnetic resonance (NMR) structural biology and NMR structural genomics pose a fascinating set of geometric challenges. A key bottleneck in NMR structural biology is the resonance assignment problem. We seek to accelerate protein NMR resonance assignment and structure determination by exploiting a priori structural information. In particular, a method known as nuclear vector replacement (NVR) has been proposed as a method for solving the assignment problem given a priori structural information. Among several different types of input data, NVR uses a particular type of NMR data known as residual dipolar couplings (RDCs). The basic physics of RDCs tells us that the data should be explainable by a structural model and set of parameters contained within the "Saupe alignment tensor".

In the NVR algorithm, one estimates the Saupe alignment tensors and then proceeds to refine those estimates. We would like to quantify the accuracy of such estimates, where we compare the estimated Saupe matrix to the correct Saupe matrix. In this work, we propose a way to quantify this comparison. Given a correct Saupe matrix and an estimated Saupe matrix, we compute an upper bound on the probability that a randomly rotated Saupe tensor would have an error smaller than the estimated Saupe matrix. This has the advantage of being a quantified upper bound, which also has a clear interpretation in terms of geometry and probability. While the specific application of our rotation probability results is given to NVR, our novel methods can be used for any RDC-based algorithm to bound the accuracy of the estimated alignment tensors. Furthermore, they could also be used in X-ray crystallography or molecular docking to quantitate the accuracy of calculated rotations of proteins, protein domains, nucleic acids, or small molecules.

KEY WORDS—$SO(3)$, rotations, subgroup method, orthogonal image, alignment tensor, residual dipolar couplings, Saupe matrix, NMR structural biology

ABBREVIATIONS USED—NMR, nuclear magnetic resonance; NVR, nuclear vector replacement; RDC, residual dipolar coupling; 3D, three-dimensional; MR, molecular replacement; SAR, structure activity relation; DOF, degrees of freedom; nt., nucleotides; $SO(3)$, special orthogonal (rotation) group in 3D; S^1, the unit circle; S^2, the unit sphere; S^3, the unit 3-sphere; FAA, frame axis angle representation of rotations.

1. Introduction

In the field of structural biology, nuclear magnetic resonance (NMR) is a powerful tool for studying the structure of proteins, as well as elucidating the interaction of proteins with other molecules. Typically, the results of protein solution-state NMR experiments yield geometric measurements such as inter-proton distances, dihedral bond angles, and global orientations of bonds. While such information is extremely useful, NMR data are initially unassigned. For example, we are typically given a protein with a known sequence of amino acids, which we simply index sequentially. NMR data will

The International Journal of Robotics Research
Vol. 24, No. 2–3, February–March 2005, pp. 165-182,
DOI: 10.1177/0278364905050351

give a set of constraints (e.g., inter-proton distances), but will reference the amino acids with a different and arbitrary indexing scheme (based on nuclear resonance frequency). The process of determining the one-to-one mapping from one indexing scheme to the other is known as “assignment”. Assignment is the solution to an inverse problem, namely, the mapping of k-tuples of resonance frequencies to the k-tuples of interacting NMR-active nuclei (Zimmerman et al. 1997; Bailey-Kellogg et al. 2000; Al-Hashimi and Patel 2002; Hus, Propmers, and Brüschweiler 2002). The assignment problem is perhaps the critical bottleneck for the interpretation and exploitation of NMR data. It is desirable to discover faster methods for solving it, as well as to exploit any formal insights about the combinatorial complexity and structure of the problem.

Recently, a number of researchers have sought to accelerate protein NMR assignment and structure determination by exploiting a priori structural information. By analogy, rapid structure determination is facilitated in X-ray crystallography by the molecular replacement (MR) technique (Rossman and Blow 1962) for solving the crystallographic phase problem. The corresponding bottleneck in NMR structural biology is the resonance assignment problem. One would hope that knowing a structural model ahead of time could expedite assignment. Moreover, even when the structure of a protein has already been determined by X-ray crystallography or computational homology modeling, NMR assignments are valuable because NMR can be used to probe: protein–protein interactions (Fiaux et al. 2002), via chemical shift mapping (Chen et al. 1993); protein–ligand binding, via structure activity relation by NMR (Shuker et al. 1996) or line-broadening analysis (Fejzo et al. 1999); and dynamics, for example via nuclear spin relaxation analysis (Palmer 1997).

To enable structure-based resonance assignment, the idea of correlating unassigned experimentally measured residual dipolar couplings (RDCs; Tjandra and Bax 1997; Losonczi et al. 1999) with bond vector orientations from a known structure was first proposed by Al-Hashimi and Patel (2002) and subsequently demonstrated in Al-Hashimi et al. (2002), who considered permutations of assignments for RNA, and Hus, Propmers, and Brüschweiler (2002), who assigned a protein from a known structure using bipartite matching. Later, we proposed a method known as nuclear vector replacement (NVR; Langmead et al. 2003, 2004; Langmead and Donald 2004), which builds on these works and offers some improvements in terms of isotopic labeling, spectrometer time, accuracy, robustness and computational complexity. Within the NVR algorithm (as well as within almost any RDC-based algorithm) it becomes necessary to interpret the NMR data known as RDCs. According to basic physics, the RDC data should be explained by the structure of the protein, as well as several parameters represented by the Saupe alignment tensor. As is well known (Losonczi et al. 1999) the alignment tensor may be represented by specifying its eigenvalues, together with a three-dimensional (3D) rotation, called the principle order frame (POF). In this paper we present a novel and rigorous method for bounding the accuracy of rotation matrices. This general method is then applied to quantitate the accuracy of POFs.

Specifically, in the NVR algorithm, one estimates the Saupe alignment tensors and then proceeds to refine those estimates. We would like to quantify the accuracy of such estimates, where we compare the estimated Saupe matrix to the correct Saupe matrix. We propose a novel way to quantify this comparison. Given a correct Saupe matrix and an estimated Saupe matrix, we compute an upper bound on the probability that a randomly-rotated Saupe tensor would have a geometric error smaller than the estimated Saupe matrix.

In Section 2, we first give a brief introduction to RDCs. Then, in Section 4.1, we explain our method for comparing Saupe alignment tensors. Finally, we present some results which quantify the accuracy of the NVR Saupe matrix estimation. While the specific application of our rotation probability results is given to NVR, these novel methods can be used for any RDC-based algorithm to bound the accuracy of the estimated alignment tensors. Furthermore, they could also be used in X-ray crystallography or molecular docking to quantitate the accuracy of calculated rotations of proteins, protein domains, nucleic acids, or small molecules.

2. Background: Brief Introduction to Residual Dipolar Couplings

RDCs are a quantum mechanical effect arising from the dipole–dipole interaction of nuclear spins. While the detailed physics are not important for our problem, we briefly explain the formalism of RDCs.

RDCs are experimentally measured real values that may be interpreted as constraints on the orientation of a chemical bond. We explain this formally; we follow Wedemeyer, Rohl, and Scheraga (2002) and Losonczi et al. (1999).

Let n be number of residues in the protein. Let $\mathbf{v}_i$ be a unit column vector in $\mathbb{R}^3$ which represents the orientation of a chemical bond $(1 \leq i \leq n)$. (We will consider only one chemical bond per residue.) Let the Saupe Matrix, $\mathbf{S}$ be a 3×3 matrix which is symmetric and traceless.

We define the RDC to be a quadratic form over the unit sphere

$$D_i = D(\mathbf{v}_i) \equiv k\,\mathbf{v}_i^{\mathrm{T}}\mathbf{S}\mathbf{v}_i, \tag{1}$$

where k is a constant based on physical constants and the dynamics of the protein in solution (Saupe 1968; Losonczi et al. 1999; Wedemeyer, Rohl, and Scheraga 2002).

Suppose we are given $\mathbf{S}$ and D_i, then eq. (1) is a constraint on the possible orientations of $\mathbf{v}_i$. In a typical RDC experiment, D_i are measured; however, both $\mathbf{v}_i$ and the Saupe matrix $\mathbf{S}$ are unknown. When computing $\mathbf{S}$, it is useful to note that $\mathbf{S}$

has only five degrees of freedom because it is real, symmetric, and traceless.

Before continuing, it is worth noting that RDCs and the Saupe matrix are of considerable interest for research in structural biology. Two of the dominant problems in NMR structural biochemistry are the assignment problem, and structure determination or refinement. As mentioned in the introduction, there is interest in using RDCs to perform structure-based resonance assignment (Tjandra and Bax 1997; Losonczi et al. 1999; Al-Hashimi and Patel 2002; Al-Hashimi et al. 2002; Hus, Propmers, and Brüschweiler 2002; Langmead et al. 2003, 2004). In addition, RDCs have been used as geometric restraints to determine and/or refine the structure of proteins (Clore, Gronenborn, and Bax 1998; Degalio, Kontaxis, and Bax 2000; Fowler et al. 2000; Andrec, Du, and Levy 2001; Tian, Valafar, and Prestegard 2001; Rohl and Baker 2002; Giesen, Homans, and Brown 2003; Wang and Donald 2004). Because of the wide range of applications of RDCs, we believe it is important to analyze and characterize the accuracy of Saupe alignment tensors.

3. Description of Problem and Previous Work

When we are given D_i and the corresponding (i.e., assigned) $\mathbf{v}_i$, we can compute the correct[1] Saupe matrix $\mathbf{S}$ via the singular value decomposition (SVD) method (Losonczi et al. 1999). Our problem, then, is to quantify the comparison between the correct Saupe matrix, and an estimated one. We observe that the Saupe matrix is real and symmetric; therefore, it has real eigenvalues and orthogonal eigenvectors. In fact, the Saupe matrices are completely specified by their eigenvalues and eigenvectors. Accordingly, our similarity measure is broken up into two parts: a comparison of eigenvalues, and then a comparison of eigenvectors.

Following standard notation (Wedemeyer, Rohl, and Scheraga 2002), we sort the eigenvectors by eigenvalue.[2] We then compare eigenvalues and eigenvectors of the same rank. For the eigenvalues, one can simply compute the relative error between the estimated and correct eigenvalues. For the eigenvectors, one can compute the angle between each correct eigenvector and its corresponding estimated eigenvector.

Both of these measures are simple and useful. The eigenvalues can be considered to have units of Hertz, which are directly comparable to the resolution of the NMR spectrum. That is to say, the resolution of the NMR spectrum gives us a length-scale for judging the accuracy of the estimated eigenvalues. For example, if the error in the eigenvalues is much larger than the resolution of the NMR spectrum, we would judge the eigenvalues to be inaccurate.

For the eigenvectors, the angular errors are simple to understand geometrically. However, it is not clear how to judge when the angular errors may be considered “small”. There are several ways to do this, such as comparing the angular errors to an angular threshold set by, say, the required accuracy for drug design, or perhaps the angular changes from protein dynamics. While many of these methods are useful, we propose a new measure of eigenvector accuracy which is purely geometrical, and contains an intuitive notion of “how difficult” it is to achieve a given angular accuracy. We see our method as a new measure which provides some additional insight, and not as a replacement of other measures.

4. Methods

4.1. Percentile Measure of Saupe Eigenvector Accuracy

We motivate our method with a simple idea: we will use probability as the judge of accuracy. Given an estimated answer, and a correct answer, we can ask whether we randomly guessed a solution. What is the probability that the random guess is closer to the correct solution than the estimated answer? Alternatively, we can ask what fraction of all possible solutions are worse than our estimated solution?

To apply this idea, we need to specify two things: first, a base measure of accuracy and, secondly, the space of all solutions. For our base measure of accuracy, we will choose the angular error between corresponding eigenvectors. Formally, we define this as follows:

$$\text{Let } \mathbf{S}_1 = \text{correct Saupe matrix} \tag{2}$$
$$\text{Let } \mathbf{S}_2 = \text{estimated Saupe matrix} \tag{3}$$
$$\text{Let } \lambda_i = \text{eigenvalues of } \mathbf{S}_1 \text{ where } \lambda_3 > \lambda_1 > \lambda_2 \tag{4}$$
$$\text{Let } \mathbf{v}_i = \text{eigenvectors of } \mathbf{S}_1 \text{ where } \mathbf{S}_1\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{5}$$
$$\text{Let } \rho_j = \text{eigenvalues of } \mathbf{S}_2 \text{ where } \rho_3 > \rho_1 > \rho_2 \tag{6}$$
$$\text{Let } \mathbf{w}_j = \text{eigenvectors of } \mathbf{S}_2 \text{ where } \mathbf{S}_2\mathbf{w}_j = \rho_j\mathbf{w}_j \tag{7}$$
$$\text{Let } \measuredangle(\mathbf{v}_i, \mathbf{w}_j) = \text{the angle between the vectors } \mathbf{v}_i \text{ and } \mathbf{w}_j \tag{8}$$
$$\text{Let } \measuredangle_{min}(\mathbf{v}_i, \mathbf{w}_j) = \min(\measuredangle(\mathbf{v}_i, \mathbf{w}_j), \measuredangle(-\mathbf{v}_i, \mathbf{w}_j)). \tag{9}$$

DEFINITION 1. Given: a Saupe matrix $\mathbf{Q}$ with eigenvalues $\gamma_3 > \gamma_1 > \gamma_2$ and corresponding eigenvectors $\mathbf{u}_k$ where $\mathbf{Q}\mathbf{u}_k = \gamma_k\mathbf{u}_k$. We say the eigenvectors of $\mathbf{Q}$ are *geometrically more accurate* than the eigenvectors of $\mathbf{S}_2$ when $\measuredangle_{min}(\mathbf{u}_i, \mathbf{v}_i) \leq \measuredangle_{min}(\mathbf{w}_i, \mathbf{v}_i)$ for all $i \in \{1, 2, 3\}$.

We define $\measuredangle_{min}$ as stated above, because we need to account for the inversion symmetry of eigenvectors. That is, if $\mathbf{v}$ is an eigenvector of $\mathbf{S}_1$ with eigenvalue λ, then so is $-\mathbf{v}$. Also, it is worth noting that we require all the corresponding eigenvectors of $\mathbf{Q}$ to have angular deviations which are smaller than the deviations of $\mathbf{S}_1$ eigenvectors.

[1] For the purposes of comparison and to quantitate the accuracy of NVR, “true” values of the alignment tensors are determined by computing the optimal Saupe matrix using the correct assignments. For this paper, it is not important how the “correct” Saupe matrix is computed.

[2] Following the convention of Wedemeyer, Rohl, and Scheraga (2002), we label the largest eigenvalue as z, the smallest as y, and the middle as x.

Next, we need to specify the space of all possible solutions. We consider all possible rotations of a Saupe matrix $\mathbf{Q}$ and its corresponding eigenvectors. Here, we encounter two issues. First, we want to consider all rotations in an isotropic manner, so that all orientations of $\mathbf{Q}$ are equally likely in an geometric sense. Secondly, we need to account for the inversion symmetry of eigenvectors: if $\mathbf{u}_j$ is an eigenvector of $\mathbf{Q}$ with eigenvalue γ_j, then so is $-\mathbf{u}_j$. This inversion symmetry will be accounted for as a multiplicative factor within our final solution. In the following two sections, we formally explain the details of how we address both issues.

4.1.1. Isotropic Representations of Rotations

There are many representations of rotations. Some examples include Euler angles, axis–angles, and quaternions. We wish to choose a representation which is isotropically uniform. Euler angles are known to have singularities in their parameterization (so-called "gimbal lock" in computer graphics). As a result, Euler angles are clearly not an isotropically uniform representation. In our work here, we choose a modified version of axis–angle, and show that it is isotropic. The motivation for our choice is simply convenience.

DEFINITION 2. Let P be a probability distribution over all possible rotations.
Let L be an arbitrary set of rotations.
Let $P(L)$ be the probability that we pick a rotation in L, if we randomly choose according to P.
Let $\mathbf{R}$ be an arbitrary rotation.
Let $\mathbf{R}L$ be the set of all rotations generated by rotating each element of L by $\mathbf{R}$.
Let $P(\mathbf{R}L)$ be the probability that we pick an rotation in $\mathbf{R}L$, if we randomly choose according to P.
We say that the probability distribution P is *rotationally symmetric* if and only if for all rotations $\mathbf{R}$, and for all possible sets of rotations L, that $P(L) = P(\mathbf{R}L)$.

For those familiar with group theory, it is worth noting that a rotationally symmetric probability distribution is a special case of the Haar measure on rotations (Diaconis and Shahshahani 1999). A Haar measure is a measure over subsets of the group, and is invariant under group operations. In our case, our group is the space of rotations, and our group operations are composition of rotations.

DEFINITION 3. Let $\mathbf{G}(\alpha, \beta, \gamma)$ be a parameterization of rotations, $SO(3)$, with parameters α, β, and γ.
Let $D(\alpha, \beta, \gamma)$ be a uniform probability distribution over the range of (α, β, γ).
Let $\mathbf{v}$ be a unit vector.
Let $F_{\mathbf{v}}$ be the distribution of unit vectors, $\mathbf{G}\mathbf{v}$, induced by D.
We say $\mathbf{G}$ is *isotropically uniform* if and only if both of the following are true:
(1) for all $\mathbf{v} \in S^2$, $F_{\mathbf{v}}$ is the uniform distribution over the unit sphere S^2;
(2) D induces a distribution over rotations that is rotationally symmetric.

Intuitively, we want a way to choose a random rotation whereby we mean that the rotation of a vector, $\mathbf{v}$, creates a new vector, $\mathbf{v}'$, which is completely randomized. In other words, the distribution of $\mathbf{v}'$ should be uniform over the sphere.

There are several ways to choose a parameterization G which is isotropically uniform. For example, there is likely to be a parameterization based on quaternions. Another example is the orthogonal image representation of rotations (Mandell et al. 2001; Mitchell 2004), which is closely related to our method (we discuss the connection below). For our purposes, we start with coordinate frames because they are easier to visualize. Frames are isomorphic to rotations, so this is simply a choice of representation (see Appendix C). We start with a modified axis–angle representation of frames and then, conceptually, we convert frames into rotations. We then show that the modified axis–angle representation of rotations is isotropically uniform. Finally, we use the geometry of the modified axis–angle representation to simplify some of the algebra when we compute our similarity measure for Saupe matrices.

Because our axis–angle representation differs from the canonical (classical) axis–angle representation of rotations, we will define both, so that they can be compared. We will call the new representation the frame-axis–angle (FAA) representation, because it is more closely related to coordinate frames. We will use the term "axis–angle" to mean the canonical (classical) representation.

DEFINITION 4. By *frame*, we mean a choice of coordinate frame. Formally, a coordinate frame is an ordered triple of unit vectors $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ such that the vectors are orthogonal to each other and oriented according to the right-hand rule, $\mathbf{x} \times \mathbf{y} = \mathbf{z}$.

DEFINITION 5. Let $\mathbf{v}$ be a unit vector in $\mathbb{R}^3$, and let $\theta \in [0, 2\pi)$.
We define the *(canonical) axis–angle representation of rotations* to be a mapping which takes $(\mathbf{v}, \theta)$ to the rotation by θ radians around the axis $\mathbf{v}$.

DEFINITION 6. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^3$, and let $\theta \in [0, 2\pi)$.
We define the *FAA representation of frames* to be a mapping which takes $(\mathbf{u}, \theta)$ to the coordinate frame specified as follows.
(1) Choose the z-axis to be along $\mathbf{u}$.
(2) Choose the x-axis to be perpendicular to $\mathbf{u}$ and rotated around $\mathbf{u}$ by an angle specified by θ. The exact position of $\theta = 0$ is arbitrary,[3] but is considered to be a constant for each choice of $\mathbf{u}$.

[3] As a result, the FAA representation is not unique, but is many parameterizations which differ only in their specification of where $\theta = 0$ for each choice of $\mathbf{u}$.

(3) Now the y-axis is uniquely determined by the requirements that the y-axis is orthogonal to the other two axes, and that the coordinate system is right-handed.

While both axis–angle and FAA use the same parameters as input (a unit vector and an angle), it should be clear that axis–angle and FAA are different. In the axis–angle representation, both the axis and angle of the rotation are explicitly specified. However, in FAA, neither the axis nor the angle of rotation is immediately obvious; the vector in axis–angle represents the axis of rotation, while in FAA it represents the new position of the z-axis.

We note that the FAA representation is a one-to-one and onto mapping between $S^2 \times S^1$ and frames. For a proof, see Appendix B. We also note that FAA is a discontinuous parameterization of frames. While this might appear to be problematic, it does not affect our proof that FAA is an isotropic representation of frames (the proof is in Appendix A). However, we do address the discontinuities explicitly when we integrate over the parameters of the FAA representation (see Section 4.1.2 and eq. (19)). For a proof that FAA is discontinuous, see Appendix B.

DEFINITION 7. Let $\mathbf{H}$ be a frame (coordinate frame) specified by $(\mathbf{x}, \mathbf{y}, \mathbf{z})$.
Let $\mathbf{v} = [x, y, z]$ be a unit vector.
We define *applying frame* $\mathbf{H}$ *to* $\mathbf{v}$ to be the vector $\mathbf{v}' = \mathbf{H}\mathbf{v} = x\mathbf{x} + y\mathbf{y} + z\mathbf{z}$.

We define isotropically uniform parameterizations of frames in a manner analogous to rotations.

DEFINITION 8. Let P be a probability distribution over all possible frames.
Let L be an arbitrary set of frames.
Let $P(L)$ be the probability that we pick a frame in L, if we randomly choose according to P.
Let $\mathbf{R}$ be an arbitrary rotation.
Let $\mathbf{R}L$ be the set of all frames generated by rotating each element of L by $\mathbf{R}$.
Let $P(\mathbf{R}L)$ be the probability that we pick a frame in $\mathbf{R}L$, if we randomly choose according to P.
We say that the probability distribution P is *rotationally symmetric* if and only if for all rotations $\mathbf{R}$, and for all possible sets of frames L, that $P(L) = P(\mathbf{R}L)$.

DEFINITION 9. Let $\mathbf{H}(\alpha, \beta, \gamma)$ be a parameterization of frames, with parameters α, β, and γ.
Let $D(\alpha, \beta, \gamma)$ be a uniform probability distribution over the range of (α, β, γ).
Let $\mathbf{v}$ be a unit vector.
Let F_v be the distribution of unit vectors, $\mathbf{H}\mathbf{v}$, induced by D.
We say $\mathbf{H}$ is *isotropically uniform* if and only if both of the following are true:
(1) for all $\mathbf{v} \in S^2$, F_v is the uniform distribution over S^2;
(2) D induces a distribution over frames that is rotationally symmetric.

We now claim that the FAA representation is isotropically uniform. The problem of how to isotropically sample rotations has been studied extensively. Our particular choice of representation is based on convenience and usefulness when applied to our particular problem (comparison of Saupe matrices). We prove that our FAA representation is isotropically uniform, by relating it to known representations of rotations that are rotationally symmetric. In particular, there is a closely related representation of rotations known as the orthogonal image representation (Mandell et al. 2001; Mitchell 2004). Our FAA representation is a direct parameterization of the orthogonal image representation. The orthogonal image representation can be seen as a special case of the subgroup algorithm (Diaconis and Shahshahani 1999). Therefore, the FAA representation is a parameterization of the subgroup algorithm. The subgroup algorithm provides a general way for computing uniformly distributed variables of compact groups.

For some intuition, and a sketch of the proof that the FAA representation is isotropically uniform, see Appendix A. For a discussion of some technical points about the FAA representation, see Appendix B. In Appendix C we discuss the relationship between a few different representations of rotations, including FAA, orthogonal image, and quaternions.

THEOREM 1. The FAA parameterization is isotropically uniform.

Proof. See Appendix A. □

FAA is an isotropic representation of frames. We now connect frames to rotations, to show that FAA is an isotropic representation of rotations. Each rotation determines a unique frame relative to the standard Euclidean frame ([1, 0, 0], [0, 1, 0], [0, 0, 1]). Similarly, each frame determines a unique rotation representing the transformation that changes the Euclidean frame into the given frame. Note that the columns of a rotation matrix are the unit vectors of its corresponding frame, and vice versa. So, this association is one-to-one and onto. In fact, frames and rotations are isomorphic (see Appendix C).

Now that we have an isotropically uniform representation of frames and rotations, we can proceed to consider orientations of Saupe matrices and their eigenvectors.

4.1.2. Orientations of Saupe Matrices and Eigenvectors

We now return to our problem of how to compare eigenvectors and orientations of Saupe matrices. We solve our problem in three stages. First, we simplify the problem by ignoring the inversion symmetry of the eigenvectors. Secondly, we approximate the solution to make the algebra more tractable. Our approximation will yield a strict upper bound on the probability. Thirdly, we account for the inversion symmetry of the eigenvectors.

Because Saupe matrices are real and symmetric, their eigenvectors are orthogonal. The orthogonal eigenvectors can be used to form a coordinate system (frame). For choosing the coordinate system, one may recall that we have sorted the eigenvectors by eigenvalues and then labeled them by their sorted rank. Using the standard labeling, we take the eigenvector with the largest eigenvalue, and label it as the z-axis; the eigenvector with the smallest eigenvalue is labeled as the y-axis; finally, the remaining eigenvector is labeled as the x-axis. For now we ignore the inversion symmetry of the eigenvectors, and assume them to be uniquely determined and to be a right-handed coordinate system. (Due to the inversion symmetry, it is always possible to choose the eigenvectors to form a right-handed coordinate system. Later, we will allow for the inversion symmetry in our calculations.)

Our now simplified problem can be stated as follows:

$$\text{Let } F_1 = (\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1) \text{ coordinate frame of correct Saupe Matrix} \tag{10}$$

$$\text{Let } F_2 = (\mathbf{x}_2, \mathbf{y}_2, \mathbf{z}_2) \text{ coordinate frame of estimated Saupe Matrix} \tag{11}$$

$$\text{Let } \measuredangle(\mathbf{v}, \mathbf{w}) = \text{the angle between the vectors } \mathbf{v} \text{ and } \mathbf{w}. \tag{12}$$

Suppose we randomly choose a new coordinate system $F_3 = (\mathbf{x}_3, \mathbf{y}_3, \mathbf{z}_3)$. We want to know the probability that F_3 is geometrically closer to F_1 than is F_2. The three constraints are:

$$\text{Let } \mathcal{C}_1 \text{ be the constraint: } \measuredangle(\mathbf{x}_1, \mathbf{x}_3) \leq \measuredangle(\mathbf{x}_1, \mathbf{x}_2) \tag{13}$$

$$\text{Let } \mathcal{C}_2 \text{ be the constraint: } \measuredangle(\mathbf{y}_1, \mathbf{y}_3) \leq \measuredangle(\mathbf{y}_1, \mathbf{y}_2) \tag{14}$$

$$\text{Let } \mathcal{C}_3 \text{ be the constraint: } \measuredangle(\mathbf{z}_1, \mathbf{z}_3) \leq \measuredangle(\mathbf{z}_1, \mathbf{z}_2) \tag{15}$$

We choose F_3 in a manner that is isotropically uniform over all frames (coordinate frames). Another way to state our problem is to ask what fraction, P_c, of all frames, F_3, satisfy $\mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$ simultaneously. By integrating characteristic functions over all frames, we can compute the fraction, $P_\mathcal{C}$ of frames that satisfy our constraints. We use the FAA representation to perform the integration in an isotropically uniform manner.

$$\text{Let } \Omega(\mathbf{v}, \theta) = \text{the frame specified by FAA (unit vector } \mathbf{v}\text{,angle } \theta). \tag{16}$$

$$\text{Let } P_\mathcal{C} = \text{the probability that } F_3 \text{ satisfies } \mathcal{C}_1, \mathcal{C}_2, \text{ and } \mathcal{C}_3. \tag{17}$$

$$\text{Let } K_i(\mathbf{v}, \theta) = \text{(characteristic function:) } K_i = 1 \text{ if } \mathcal{C}_i \text{ is satisfied for } F_3 = \Omega(\mathbf{v}, \theta)\text{, otherwise, } K_i = 0. \tag{18}$$

$$\begin{aligned} \text{Let } P_\mathcal{C} &= Prob(\mathcal{C}_1 \wedge \mathcal{C}_2 \wedge \mathcal{C}_3) \\ &= \frac{1}{4\pi} \int\limits_{S^2} \frac{1}{2\pi} \int\limits_{[0,2\pi]} K_1(\mathbf{v}, \theta) K_2(\mathbf{v}, \theta) K_3(\mathbf{v}, \theta) \, \mathrm{d}\theta \, \mathrm{d}A. \end{aligned} \tag{19}$$

The integral (19) is over the unit sphere $\mathbf{v} \in S^2$ with area element $\mathrm{d}A$, and the unit circle $\theta \in S^1$ with line element $\mathrm{d}\theta$. We parametrize the unit circle by the angle $\theta \in [0, 2\pi]$.

At this point, we need to address a technical point, namely the fact that FAA is not a continuous representation of frames (see Appendix B). Although the FAA parameterization of frames is not continuous, the set of all frames is continuous. Conceptually, we are integrating over all frames, but we are forced to parametrize the set to simplify the computation. The discontinuities will not affect our integral, eq. (19), as long as the integrand is bounded in the neighborhood of the discontinuities, and the discontinuities occur on a set of measure zero. The integrand is bounded because the characteristic functions $K_i(\mathbf{v}, \theta)$ are bounded.

To ensure that the discontinuities are a set of measure zero, we consider a special case of FAA. As mentioned in the footnote for Definition 6, the FAA representation is not unique. To completely specify an FAA representation, $(\mathbf{u}, \theta)$, we need to define the new location of the x-axis when $\theta = 0$ for each $\mathbf{u}$. We choose a unique FAA representation which has discontinuities on a set of measure zero. Let $R(\mathbf{u})$ be the rotation that maps the z-axis to $\mathbf{u}$ by moving the z-axis along the geodesic between them. Next, let $R_\mathbf{u}(\theta)$ be a rotation by θ degrees around the axis $\mathbf{u}$. We can now uniquely choose our representation as $FAA(\mathbf{u}, \theta) = R_\mathbf{u}(\theta) R(\mathbf{u})$. This mapping from $(\mathbf{u}, \theta)$ to rotations (frames) is continuous everywhere except when the geodesic between $\mathbf{u}$ and the z-axis is not uniquely defined. On the 2-sphere, geodesics are unique except between antipodal points. As a result, our mapping is continuous everywhere except when $\mathbf{u}$ points in the negative z-direction. For our FAA parameterization, the discontinuity at $\mathbf{u} = [0, 0, -1]$ occurs only on a set of measure zero, therefore our integral (19) is unaffected by the discontinuity.

Now, computing $P_\mathcal{C}$ exactly is likely to be very complicated algebraically. Instead, we simplify the algebra by computing an upper bound on $P_\mathcal{C}$. We obtain an upper bound by replacing K_1 and K_2 with the upper bound on their individual values. First, we replace the factor K_2 with unity. This is equivalent to relaxing away our constraint $\mathcal{C}_2$.

Secondly, we will find an upper bound on K_1. However, before we obtain an upper bound on K_1, we wish to simplify the integral slightly. Notice that K_3 corresponds to constraint $\mathcal{C}_3$ which constrains only the z-axis of F_3. We see that K_3 has no dependence on θ. (The choice of θ only rotates about the axis $\mathbf{v} = \mathbf{z}_3$, and does not change the direction of $\mathbf{v}$.) As a result, we can pull K_3 outside the innermost integral. We are left with

$$K_3(\mathbf{v}, \theta) = K_3(\mathbf{v}) \tag{20}$$

$$P_\mathcal{C} \leq \frac{1}{8\pi^2} \int\limits_{S^2} K_3(\mathbf{v}) \int\limits_{[0,2\pi]} K_1(\mathbf{v}, \theta) \, \mathrm{d}\theta \, \mathrm{d}A. \tag{21}$$

To obtain an upper bound on K_1, we consider the geometry of the problem. The inner integral of K_1 can be thought of as a function of $\mathbf{v}$. We can simply ask, then, what is the maximum value of that function, over all possible $\mathbf{v}$?

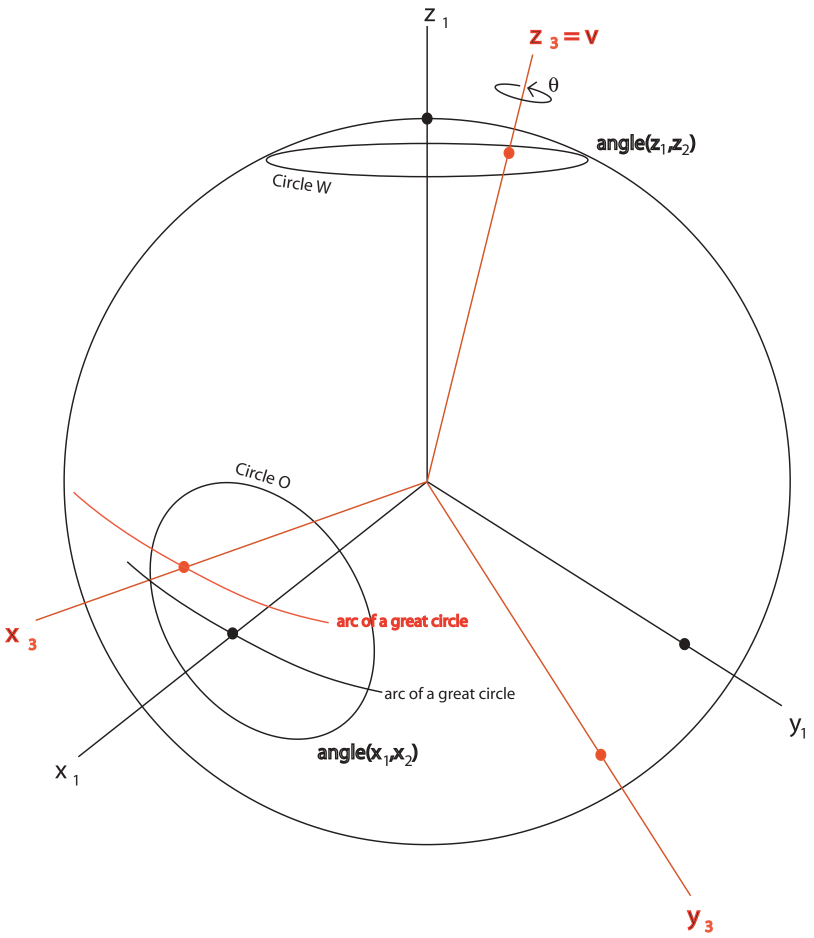

Fig. 1. Diagram for upper bound on the integral of K_1. We start with the unit sphere. We consider two coordinate frames $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1)$ and $(\mathbf{x}_3, \mathbf{y}_3, \mathbf{z}_3)$. The two circles are labeled by their angular radii $\measuredangle(\mathbf{x}_1, \mathbf{x}_2)$ and $\measuredangle(\mathbf{z}_1, \mathbf{z}_2)$. These two circles represent our constraints $\mathcal{C}_1$ and $\mathcal{C}_3$ respectively (see eqs. (13)–(15)). The circle around $\mathbf{x_1}$ we call O, and we call the other circle W. The constraint $\mathcal{C}_1$ requires that $\mathbf{x}_3$ fall inside of the circle O, and the constraint $\mathcal{C}_3$ requires that $\mathbf{z}_3$ fall inside of the circle W. Notice that once $\mathbf{z}_3$ is fixed, $\mathbf{x}_3$ can travel along a great circle, as θ rotates $\mathbf{x}_3$ around $\mathbf{v}$ (by construction $\mathbf{v} = \mathbf{z}_3$). The maximal range of θ which satisfies our constraint $\mathcal{C}_1$ cannot be larger than the diameter of the circle O which is $2\measuredangle(\mathbf{x}_1, \mathbf{x}_2)$.

$$\text{Let } J_1 = \max_{\mathbf{v}\in S^2} \int_{[0,2\pi]} K_1(\mathbf{v}, \theta)\, \mathrm{d}\theta. \tag{22}$$

To visualize the geometry of our problem, let us work in the coordinate frame F_1 of the correct Saupe matrix (see Figure 1). Our constraint $\mathcal{C}_1$ can be visualized as a small circle sitting on the equator (in the figure, this circle is labeled O and has an angular radius of $\measuredangle(\mathbf{x}_1, \mathbf{x}_2)$). The $\mathcal{C}_1$ requirement that $\measuredangle(\mathbf{x}_1, \mathbf{x}_3) \le \measuredangle(\mathbf{x}_1, \mathbf{x}_2)$ means that we require $\mathbf{x}_3$ to fall within this circle. For a fixed $\mathbf{v}$, we have $\mathbf{z}_3 = \mathbf{v}$ is also fixed. Then, as θ rotates around, the location of $\mathbf{x}_3$ will sweep out a great circle. If this great circle intersects the circle O, then we can satisfy our condition $\mathcal{C}_1$. If the great circle does not intersect the circle O, then for this choice of $\mathbf{v}$, there is no angle θ which can satisfy $\mathcal{C}_1$.

Let us return to eq. (22). For a fixed choice of $\mathbf{v}$, the value of the integral is equal to the angle for which the indicator function $K_1(\mathbf{v}, \theta)$ is equal to unity (non-zero). This angle is simply the range of θ for which $\mathbf{x}_3$ falls inside the circle. This is simply the length of the arc, of the great circle of $\mathbf{x}_3$, which is inside the circle O. Over all possible choices of $\mathbf{v}$, the maximal length of this arc cannot be more than the angular diameter of the circle O. Therefore, an upper bound on eq. (22) is simply

$$2\measuredangle(\mathbf{x}_1, \mathbf{x}_2) \ge J_1 = \max_{\mathbf{v}\in S^2} \int_{[0,2\pi]} K_1(\mathbf{v}, \theta)\, \mathrm{d}\theta. \tag{23}$$

Our equation for satisfying our constraints now looks like

$$P_{\mathcal{C}} \leq \frac{1}{8\pi^2} \int_{S^2} 2K_3(\mathbf{v}) \measuredangle(\mathbf{x}_1, \mathbf{x}_2)\, \mathrm{d}A \tag{24}$$

$$= \frac{\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{4\pi^2} \int_{S^2} K_3(\mathbf{v})\, \mathrm{d}A. \tag{25}$$

The final integral involving K_3 is simply the area of the disc where $\mathbf{v}$ satisfies $\measuredangle(\mathbf{z}_1, \mathbf{v}) \leq \measuredangle(\mathbf{z}_1, \mathbf{z}_2)$. We can perform this integral in polar coordinates to obtain

$$\int_{S^2} K_3(\mathbf{v})\, \mathrm{d}A = \int_0^{\measuredangle(\mathbf{z}_1,\mathbf{z}_2)} \int_0^{2\pi} \sin(\theta)\, \mathrm{d}\phi\, \mathrm{d}\theta$$

$$= 2\pi(-\cos(\measuredangle(\mathbf{z}_1, \mathbf{z}_2)) + \cos(0)). \tag{26}$$

Combining our results, we obtain

$$P_{\mathcal{C}} \leq \frac{\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{4\pi^2} \int_{S^2} K_3(\mathbf{v})\, \mathrm{d}A \tag{27}$$

$$P_{\mathcal{C}} \leq \frac{\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{2\pi} (1 - \cos(\measuredangle(\mathbf{z}_1, \mathbf{z}_2))) \tag{28}$$

$$P_{\mathcal{C}} \leq \frac{\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{2\pi} (1 - \mathbf{z}_1 \cdot \mathbf{z}_2). \tag{29}$$

4.2. Inversion Symmetries and Eigenvectors

We now account for the effects of inversion symmetry on $P_{\mathcal{C}}$. Originally, we assumed that the eigenvectors were uniquely determined, and were oriented as a right-handed coordinate system. While orthogonal, the eigenvectors are not uniquely determined due to an inversion symmetry. An eigenvector $\mathbf{v}$ has the same eigenvalue as $-\mathbf{v}$. Thus, both $\mathbf{v}$ and $-\mathbf{v}$ are possible eigenvectors for a given eigenvalue.

To account for this, we shall use the idea of measuring the angle between lines which pass through the origin. These lines are similar to vectors, however, they are "bi-directional" in the sense that they do not have a definite direction like a vector. Our characterization then is to replace our eigenvectors with lines parallel to the eigenvectors, and which pass through the origin.

DEFINITION 10. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be vectors.
Let l_1 and l_2 be the corresponding lines of $\mathbf{v}_1$ and $\mathbf{v}_2$
We define *the angle between* l_1 *and* l_2 to be equal to $\measuredangle_{min}(\mathbf{v}_1, \mathbf{v}_2) = \min(\measuredangle(\mathbf{v}_1, \mathbf{v}_2), \measuredangle(\mathbf{v}_1, -\mathbf{v}_2))$.

Note that this definition measures the smaller angle between two intersecting lines. We choose the smaller angle, so that identical lines will have an angular difference of zero. In addition, this approach will overestimate the probability $P_{\mathcal{C}}$, because it may believe that vectors which are nearly π radians off are very close. Thus, it will tend to include extra frames in the overestimate. This is consistent with our approach of computing an upper bound on $P_{\mathcal{C}}$.

Notice that in Definition 10 we need only consider two cases: $(\mathbf{v}_1, \mathbf{v}_2)$ and $(\mathbf{v}_1, -\mathbf{v}_2)$. This is significant, because it means that for each angle we constrain, we have only two possibilities. Our angular constraints (13)–(15) now become

$$\text{Let } \mathcal{C}_1' \text{ be the constraint: } \measuredangle_{min}(\mathbf{x}_1, \mathbf{x}_3) \leq \measuredangle_{min}(\mathbf{x}_1, \mathbf{x}_2) \tag{30}$$

$$\text{Let } \mathcal{C}_2' \text{ be the constraint: } \measuredangle_{min}(\mathbf{y}_1, \mathbf{y}_3) \leq \measuredangle_{min}(\mathbf{y}_1, \mathbf{y}_2) \tag{31}$$

$$\text{Let } \mathcal{C}_3' \text{ be the constraint: } \measuredangle_{min}(\mathbf{z}_1, \mathbf{z}_3) \leq \measuredangle_{min}(\mathbf{z}_1, \mathbf{z}_2). \tag{32}$$

If we were to perform the same analysis as above, we discover that we gain a factor of 2, for two out of the three constraints. Why only two out of three? Because a rotation is completely specified by the mapping of two vectors. Once $\mathbf{z}_3$ and $\mathbf{x}_3$ are specified, the line representing the y-axis is fixed. The only possibilities are that $\mathbf{y}_3$ satisfies $\mathcal{C}_2'$ or not, and there are not multiple ways to satisfy or violate the condition.

Conceptually, we first satisfy $\mathcal{C}_3'$. For each choice of $\mathbf{z}_3$, $\measuredangle_{min}$ forces us to look at the choice $-\mathbf{z}_3$. These two possibilities represent two rotations that satisfy the constraint $\mathcal{C}_3'$. Next, we attempt to satisfy $\mathcal{C}_1'$. For each choice of $-\mathbf{z}_3$, and for each choice of $\mathbf{x}_3$ that satisfies $\mathcal{C}_1'$, we know that a choice of $-\mathbf{x}_3$ which also satisfy $\mathcal{C}_1'$. Combining these possibilities, we obtain a total of four possible solutions which satisfy $\mathcal{C}_1'$ and $\mathcal{C}_3'$.

At this point, one might wonder if the inversion symmetry can also be applied to $\mathcal{C}_2'$ to generate a total of eight solutions. However, this is impossible. If we invert $\mathbf{y}_3$ to become $-\mathbf{y}_3$, then our frame may no longer be valid (generated by a pure rotation). Instead, we may generate what is known as a "perversion", which is a pure rotation composed with a planar reflection. (Perversions change the handedness of our coordinate system.) Our stated problem only considers all possible rotations and does not include perversions.

Stated a different way, the choices of inverting (or not inverting) $\mathbf{z}_3$ and $\mathbf{x}_3$ can be accounted for by modifying the parameters $\mathbf{v}$ and θ in the FAA representation. However, once $\mathbf{z}_3$ and $\mathbf{x}_3$ are specified, one is not free to choose $\mathbf{y}_3$ because it is fully determined. Independently specifying the inversion (or non-inversion) of $\mathbf{y}_3$ cannot be accounted for by $\mathbf{v}$ and θ, because the resulting transformation may not be a rotation (frame). Similarly, one is free to specify any two out of the three axes as possibilities for inversion. However, one cannot choose all three.

As a result, we need to modify our eq. (29) for $P_{\mathcal{C}}$ by a factor of 4:

$$P_{\mathcal{C}} \leq \frac{2\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{\pi} (1 - \mathbf{z}_1 \cdot \mathbf{z}_2). \tag{33}$$

Finally, we convert this probability into a lower bound on the percentile:

$$\mathit{Percentile} = 1 - P_{\mathcal{C}} \tag{34}$$

$$\mathit{Percentile} \geq 1 - \frac{2\measuredangle(\mathbf{x}_1, \mathbf{x}_2)}{\pi} (1 - \mathbf{z}_1 \cdot \mathbf{z}_2). \tag{35}$$

This percentile represents the fraction of all frames that have a greater geometric difference from F_1 than the geometric difference between F_1 and F_2.

5. Applications to NMR Residual Dipolar Couplings

We use our percentile measure, e.g., (35) to characterize and quantify the accuracy of estimates of Saupe matrices. In particular, we have investigated the accuracy of Saupe matrices in the NVR algorithm (Langmead et al. 2003, 2004; Langmead and Donald 2004). Briefly, the NVR algorithm is designed to solve the NMR assignment problem when the structure of the target protein is known, or if a homologous structure is known. To achieve this, NVR uses a variety of data, including a model structure, RDCs, an HSQC spectrum, amide exchange data, and unassigned NOEs. To correlate the RDC data against a model structure, NVR needs an estimate of the Saupe matrices. Given the Saupe matrix and an NH bond vector, we can use eq. (1) to convert the vector into a simulated RDC value. The simulated RDC value can then be compared to experimentally observed RDC values during the assignment process. We refer the reader to Langmead et al. (2003, 2004) for the details of the NVR algorithm.

The NVR algorithm was demonstrated on NMR data from a 76-residue protein, human ubiquitin, matched to four structures, including one mutant (homolog), determined either by X-ray crystallography or by different NMR experiments (without RDCs; Langmead et al. 2003, 2004). The feasibility of NVR was further demonstrated for different and larger proteins, using different combinations of real and simulated NMR data for hen lysozyme (129 residues) and streptococcal protein G (56 residues), matched to a variety of 3D structural models (Langmead et al. 2003, 2004); see Table 7.

The first stage of NVR is to estimate the Saupe matrices. Sufficient accuracy in this first stage is important, since subsequent stages of the algorithm depend on a reasonable initial estimate. The initial estimate is usually performed by employing a small number (around five) of high-confidence assignments. The resulting Saupe matrix is then used to refine probabilities of assignments, and then additional assignments are made. The additional assignments are used to refine the Saupe matrix further, and the process is repeated until assignment is complete. (For details of the NVR algorithm, see Langmead et al. 2003, 2004). Because of its iterative nature and lack of back-tracking, a poor initial estimate of the Saupe matrix could lead to additional assignments that are incorrect. The accuracy of the Saupe matrix might be degraded if it is “refined” with these incorrect assignments. It is then possible that the entire iterative cycle would diverge from the correct assignment. As a result, we are interested in quantifying the accuracy of the initial and final Saupe matrices in NVR.

We characterize the accuracy of the Saupe matrix estimates in three ways. First, we compare the eigenvalues by looking at percentage differences of the axial and rhombic components of the tensor (Wedemeyer, Rohl, and Scheraga 2002):

$$\text{Let } \mathbf{S} = \text{a Saupe matrix} \tag{36}$$

$$\text{Let } \lambda_i = \text{eigenvalues of } \mathbf{S} \text{ where } \lambda_3 > \lambda_1 > \lambda_2. \tag{37}$$

Definition 11. The *axial component* of **S** is defined to be $D_a = (1/2)\lambda_3$.

Definition 12. The *rhombic component* of **S** is defined to be $D_r = (1/3)(\lambda_1 - \lambda_2)$.

Next, we consider the angular error between corresponding eigenvectors. Finally, we use our percentile measure to characterize the fraction of all orientations which have larger angular errors. See Tables 1–3. As we can see, the percentiles are above 80%, with typical values above 94%. This level of accuracy is sufficient for subsequent stages in NVR to achieve good accuracy for assignment (see Table 7). After assignment is complete, the Saupe matrices have been refined with very good accuracies (see Tables 4–6). We note that when NVR completes an assignment with very high accuracy, the corresponding Saupe tensor will be very close to the actual tensor. This is because the correct tensor is the Saupe matrix that optimally fits the protein structure given the correct assignment (see footnote 1).

Before assignment, a few of the angular deviations appear to be significantly large (approaching 30°; see Tables 1–3); however, the percentile measure shows the difference in overall rotation to be small (percentiles over 94%). Despite apparently significant angular deviations, NVR converges to assignments with high accuracy. This may suggest that for our specific case here (NVR), the percentile measure could be more useful than angular deviations for characterizing the accuracy of Saupe matrices, in the sense that it might be a more accurate indicator of when NVR will converge with high accuracy. We believe these results indicate that the percentile measure has potential to provide some insight for many applications of RDCs.

6. Conclusions

We have presented a novel similarity measure for quantifying the error of eigenvectors of Saupe matrices. This was done by developing a probability-based similarity measure for 3D rotations. The similarity measure yields a lower bound of a percentile, which represents the probability that a randomly rotated Saupe matrix would contain eigenvectors that have a larger angular deviation. We then used this percentile measure to study the performance of the automated NMR assignment method NVR (Langmead et al. 2003, 2004). We believe that the percentile measure will be useful in quantifying the performance of many NMR algorithms which utilize RDCs. In addition, our ideas may also help elucidate the performance of

Table 1. Ubiquitin Tensor Estimates

	Bicelle 292 K						Bicelle 298 K					
	Percent difference		Angular Difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
1G6J	2.3	0.2	20.8	25.1	21.8	98	12.0	5.0	28.1	30.3	16.1	96
1UBI	1.1	3.7	27.3	28.2	7.1	96	15.2	8.3	28.4	17.8	27.7	96
1UBQ	0.8	2.6	17.5	11.7	20.8	99	15.3	7.9	16.4	27.3	32.0	95
1UD7	0.2	2.2	21.2	16.5	25.8	98	14.7	6.9	16.9	16.3	7.4	99

This table demonstrates the accuracy of the first step of the NVR algorithm: tensor estimation. Columns 2 and 3 show the percentage difference for the axial and rhombic terms, D_a and D_r, for the four models, 1G6J, 1UBI, 1UBQ and 1UD7, versus the actual axial and rhombic terms in the bicelle medium recorded at 292 K. The D_a and D_r differences are normalized by the range of the experimentally measured dipolar coupling values. Columns 4–6 show the angular differences (in degrees) between the eigenvectors of the estimated tensors and the eigenvectors of the actual tensors in the bicelle medium at 292 K. S_{zz} is the director of the tensor (i.e., the eigenvector associated with the largest eigenvalue of the tensor). S_{xx} and S_{yy} are eigenvectors associated with the second largest and smallest eigenvalue of the tensor, respectively. Columns 8–12 show the accuracy of the tensor estimates in the bicelle medium recorded at 298 K. Columns 7 and 13 report the accuracy of the tensor estimate as a percentile (eq. (34)).

Table 2. Streptococcal Protein G (SPG) Tensor Estimates. Tensor estimates for the B1 domain of SPG

	Phage						Bicelle					
	Percent difference		Angular difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
1GB1	0.6	6.0	26.8	23.3	21.4	97	2.4	6.6	17.9	20.5	22.3	98
2GB1	0.2	0.5	26.8	23.3	21.4	97	1.7	10.3	17.9	20.5	22.3	98
1PGB	0.6	6.0	23.8	24.5	28.8	97	2.4	6.6	15.2	29.3	25.8	96

Table 3. Lysozyme Tensor Estimates

	5% Bicelle						7.5% Bicelle					
	Percent difference		Angular difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
193L	1.5	0.1	16.7	6.7	16.7	99	8.8	8.7	38.6	49.0	33.2	85
1AKI	2.3	0.5	13.2	10.6	8.5	99	10.0	9.3	23.2	51.0	45.2	81
1AZF	1.7	0.5	7.6	7.3	5.6	99	9.5	8.5	31.2	29.6	11.0	95
1BGI	1.2	0.7	30.0	8.5	29.8	96	8.9	9.4	24.6	43.8	35.7	89
1H87	2.1	0.2	26.2	29.9	34.2	94	9.9	8.6	23.8	15.3	25.8	97
1LSC	1.7	0.4	16.1	20.8	22.8	98	8.9	8.5	12.2	12.0	11.6	99
1LSE	1.7	0.4	12.6	49.2	44.5	83	9.5	8.3	29.2	48.2	42.1	84
1LYZ	9.8	5.0	10.7	21.4	18.5	99	18.9	8.5	21.3	21.0	24.1	98
2LYZ	3.5	1.8	20.8	16.2	16.2	99	11.56	8.3	23.8	25.0	7.5	98
3LYZ	4.3	2.4	20.0	31.4	25.2	96	12.7	8.0	27.8	38.1	4.4	96
4LYZ	3.1	2.3	24.0	9.3	24.0	98	12.6	8.6	12.7	14.5	17.7	99
5LYZ	3.1	2.3	23.9	9.3	24.0	98	12.6	8.6	12.7	14.5	17.7	99
6LYZ	3.0	0.7	15.7	16.8	16.8	99	11.0	8.6	26.6	37.3	46.0	87

other rotation-based algorithms in structural biology, computational chemistry, and drug design, by quantifying the error of orientations and rotations of chemical bonds, domains, proteins, and ligands.

In closing, we make an observation about our final result, and consider future possibilities for investigation. Suppose that the errors in angles between two Saupe matrices are roughly equal. That is, suppose $\alpha = \measuredangle(\mathbf{x}_1, \mathbf{x}_2) = \measuredangle(\mathbf{z}_1, \mathbf{z}_2)$. In this case, we note that a Taylor expansion of eq. (34) is $O(\alpha^3)$ for small angles alpha. As a result, our percentile (35) converges very rapidly to unity, when the angular errors become small.

Finally, we note there are other approaches to comparing Saupe matrices that are likely to be useful. One approach is to assume a uniform distribution of chemical bond orientations, and then to compare the distribution of RDC values generated by each Saupe matrix. For example, one could imagine performing a simple RMSD comparison, or a more sophisticated Hausdorff-based comparison (Huttenlocher and Kedem 1990; Donald, Kapur, and Mundy 1992). Even beyond that, there

Table 4. Ubiquitin Tensor Improvements

	Bicelle 292 K						Bicelle 298 K					
	Percent difference		Angular difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
1G6J	0	0.6	0.2	0.3	0.2	100	0	0	0.2	0.2	0.1	100
1UBI	0.1	0.2	2.3	2.4	0.6	100	0	0	0.2	0.2	0.1	100
1UBQ	0	0	0	0	0	100	0	0	0	0	0	100
1UD7	0	0.1	0.5	0.2	0.5	100	0	0	0.7	0.9	0.6	100

The accuracies of the final tensor estimates, after NVR has completed the resonance assignment phase. The accuracy is improved from the initial tensor estimates (see Table 1).

Table 5. SPG Tensor Improvements

	Phage						Bicelle					
	Percent difference		Angular difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
1GB1	0	0	0	0	0	100	0	0	0	0	0	100
2GB1	0	0	0	0	0	100	0	0	0	0	0	100
1PGB	0	0	0	0	0	100	0	0	0	0	0	100

The accuracies of the final tensor estimates, after NVR has completed the resonance assignment phase. The accuracy is improved from the initial tensor estimates (see Table 2).

Table 6. Lysozyme Tensor Improvements

	5% Bicelle						7.5% Bicelle					
	Percent difference		Angular difference				Percent difference		Angular difference			
Model	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile	D_a	D_r	S_{zz}	S_{xx}	S_{yy}	Percentile
193L	0	0	0	0	0	100	0	0	0	0	0	100
1AKI	0	0	0	0	0	100	0	0	0	0	0	100
1AZF	0	0	0	0	0	100	0	0	0	0	0	100
1BGI	0	0	0	0	0	100	0	0	0	0	0	100
1H87	0	0	0	0	0	100	0	0	0	0	0	100
1LSC	0.1	0.1	0	0.1	0.1	100	0	0.1	0	0	0	100
1LSE	0	0	0	0	0	100	0	0	0	0	0	100
1LYZ	0	0	0	0	0	100	0	0	0	0	0	100
2LYZ	0	0	0	0	0	100	0	0	0	0	0	100
3LYZ	0	0	0	0	0	100	0	0	0	0	0	100
4LYZ	0	0	0	0	0	100	0	0	0	0	0	100
5LYZ	0	0	0	0	0	100	0	0	0	0	0	100
6LYZ	1.5	3.3	0.7	1.2	1.0	100	1.9	5.8	0.8	5.3	5.2	100

The accuracies of the final tensor estimates, after NVR has completed the resonance assignment phase. The accuracy is improved from the initial tensor estimates (see Table 3).

may be comparison methods that are based on the geometry of the protein in question, and may include physical effects such as flexibility and dynamics. It may then be possible for a probability measure to be defined for comparing eigenvalues in addition to our method for comparing eigenvectors.

Appendix A: Proof that the Frame-Axis–Angle Representation is Isotropically Uniform

In this appendix, we discuss and prove that the FAA representation is isotropically uniform. Before we begin, it is worthwhile to describe some incorrect intuition. As noted by Diaconis and Shahshahani (1999) and Mitchell (2004), a common error in choosing an isotropically uniform parameterization of rotations is to use the conventional (canonical) axis–angle representation of rotations (Definition 5). For the axis–angle, a uniform distribution over the parameters **v** over the sphere S^2, and the unit circle $\theta \in [0, 2\pi)$, will not induce a rotationally symmetric distribution over rotations. Superficially, this geometric construction appears to be rotationally symmetric; however, it does not respect the detailed group structure of rotations. As noted by Kendall and Moran (1963) and Diaconis and Shahshahani (1999), the distribution over θ should not be uniform, but in fact, proportional to $\sin^2(\theta)$.

We begin with a brief discussion about the intuition behind the FAA representation. The intuitive motivation is that a rotationally isotropic parameterization must, by definition,

Table 7. Accuracy

PDB ID	Exp. Method	Accuracy
1G6J (Babu, Flynn, and Wand 2001)	NMR	97
1UBI (Ramage et al. 1994)	X-ray (1.8 Å)	92
1UBQ (Vijay-Kumar, Bugg, and Cook 1987)	X-ray (1.8 Å)	100
1UD7 (Johnson et al. 1999)	NMR	93

(A) Ubiquitin

PDB ID	Exp. Method	**Accuracy**
1GB1 (Gronenborn et al. 1991)	NMR	100%
2GB1 (Gronenborn et al. 1991)	NMR	100%
1PGB (Gallagher et al. 1994)	X-ray (1.92 Å)	100%

(B) SPG

PDB ID	Exp. Method	**Accuracy**
193L (Vaney et al. 1996)	X-ray (1.3 Å)	100%
1AKI (Artymiuk et al. 1982)	X-ray (1.5 Å)	100%
1AZF (Lim et al. 1998)	X-ray (1.8 Å)	100%
1BGI (Oki et al. 1999)	X-ray (1.7 Å)	100%
1H87 (Girard et al. 2001)	X-ray (1.7 Å)	100%
1LSC (Kurinov and Harrison 1995)	X-ray (1.7 Å)	98%
1LSE (Kurinov and Harrison 1995)	X-ray (1.7 Å)	100%

(C) Lysozyme

PDB ID	Exp. Method	**Accuracy**
1LYZ (Diamond 1974)	X-ray (2.0 Å)	100%
2LYZ (Diamond 1974)	X-ray (2.0 Å)	100%
3LYZ (Diamond 1974)	X-ray (2.0 Å)	100%
4LYZ (Diamond 1974)	X-ray (2.0 Å)	100%
5LYZ (Diamond 1974)	X-ray (2.0 Å)	100%
6LYZ (Diamond 1974)	X-ray (2.0 Å)	97%

(D) Lysozyme (continued)

(A) NVR achieves an accuracy of 92–100% on the four ubiquitin models. The structure 1D3Z (Cornilescu et al. 1998) is the only published structure of ubiquitin to have been refined against RDCs. The RDCs used in Cornilescu et al. (1998) have been published and were used in each of the four NVR trials. 1G6J, 1UBI, and 1UBQ have 100% sequence identity to 1D3Z. 1UD7 is a mutant form of human ubiquitin. As such, it demonstrates the effectiveness of NVR when the model is a close homolog of the target protein. (B)–(D) The RDCs for the B1 domain of SPG (Kuszewski, Gronenborn, and Clore 1999) and hen lysozyme (Schwalbe et al. 2001) were obtained from the PDB. NOEs and amide exchange data were extracted from their associated restraints files. NVR achieves an accuracy of 100% (Table 7B) and 97–100% (Tables 7C and 7D), respectively.

place the z-axis uniformly over the unit sphere. Then, for a given placement of the z-axis, the x-axis must be uniformly distributed in a unit circle perpendicular to the new position of the z-axis. The idea for this intuition is a symmetry argument. If the distribution of x is not uniform over the circle, it seems unlikely that the parameterization is isotropically uniform, because it seems not to be rotationally symmetric about the new position of the z-axis. While this intuition is helpful, it is obviously not a proof.

In the proof, we will proceed in two steps. First, we show that a uniform probability distribution over the FAA parameters induces a distribution over frames which is rotationally symmetric. To prove this, we note that the FAA representation is a uniform parameterization of the subgroup method (Diaconis and Shahshahani 1999). Secondly, we show that a rotationally symmetric distribution of frames will randomize any unit vector so that it (the vector) becomes uniformly distributed over the unit sphere. Together, these two steps show that the FAA representation of rotations is isotropically uniform.

Lemma 1. Consider the FAA parameterization of orientations. Let $(\mathbf{v}, \theta)$ be the variables of the parameterization, where $\mathbf{v}$ is a unit vector, and θ is an angle in the range $[0, 2\pi)$. Let $P_1(\mathbf{v})$ be a uniform distribution over the unit sphere S^2. Let $P_2(\theta)$ be a uniform distribution over the unit circle S^1 such that $\theta \in [0, 2\pi)$.

Let $P(\mathbf{v}, \theta) = P_1(\mathbf{v})P_2(\theta)$ be the probability distribution which is uniform over the parameters of FAA.

We claim that $P(\mathbf{v}, \theta)$ induces a distribution over orientations $(\mathbf{v}, \theta)$ which is rotationally symmetric.

Proof. According to the subgroup method (Diaconis and Shahshahani 1999), a rotationally symmetric probability distribution over rotations may be chosen as follows. First, perform a random rotation about the z-axis which is uniform over all possible angles in $[0, 2\pi)$. Secondly, rotate the z-axis to a random point on the unit sphere, in such a way that the z-axis is uniformly distributed over the sphere. For details and the proof, we refer the reader to Diaconis and Shahshahani (1999). The subgroup method generates a probability distribution over rotations, which in turn induces a probability distribution over the parameters of the FAA representation. We argue that the induced distribution is uniform over the FAA parameters.

Let $R_z(\theta) =$ the rotation around the z-axis by θ degrees.

Let $R(\mathbf{n}) =$ the rotation, as specified by the subgroup method, which rotates the z-axis into the unit vector $\mathbf{n}$. For our proof here, the exact details of $R(\mathbf{n})$ are unimportant, except for the fact that $R(\mathbf{n})$ is a fixed function of $\mathbf{n}$.

Let $R(\theta, \mathbf{n}) = R(\mathbf{n})R_z(\theta)$ be the rotation represented by the subgroup method.

According to the subgroup method, a choice of **n** which is uniform over the unit sphere, and a choice of θ which is uniform over $[0, 2\pi)$ induces a rotationally symmetric distribution over rotations (in the language of group theory, it induces a probability distribution which respects the Haar measure).

Notice that in the subgroup method, the rotation of angle θ is performed first, and then the placement of the z-axis along **n** is performed afterwards. This is the reverse order from our FAA representation, where we first place the z-axis along **v** first, and then rotate by θ degrees around **v** afterwards. In general, rotations do not commute. What we would like to do, then, is to rewrite the subgroup rotation $R(\theta, \mathbf{n}) = R(\mathbf{n})R_z(\theta)$ in a form where $R(\mathbf{n})$ occurs first, and then is followed by a new rotation which takes the place of $R_z(\theta)$.

Consider a rotation about **n** by θ degrees. Given that $R(\mathbf{n})$ rotates the z-axis to **n**, we can think of $R(\mathbf{n})$ as a change of basis. Using the change of basis, we know the following.

$$\text{Let } R_{\mathbf{n}}(\theta) = \text{A rotation around the axis } \mathbf{n} \text{ by } \theta \text{ degrees.} \tag{38}$$

$$R_{\mathbf{n}}(\theta) = R(\mathbf{n})R_z(\theta)R^{-1}(\mathbf{n}) \tag{39}$$

$$R_{\mathbf{n}}(\theta)R(\mathbf{n}) = R(\mathbf{n})R_z(\theta) \tag{40}$$

$$R_{\mathbf{n}}(\theta)R(\mathbf{n}) = R(\theta, \mathbf{n}). \tag{41}$$

We now see that the subgroup method can also be thought of as placing the z-axis along **n** first, and then performing a rotation about **n** by θ degrees. This is very close to the definition of the of the FAA representation (Definition 6); however, it differs in that θ in FAA refers to an absolute orientation, while θ in $R_{\mathbf{n}}(\theta)$ refers to a rotation.

Given an arbitrary pair $(\theta_1, \mathbf{n})$, consider the images of the x- and z-axis after the rotation $R(\theta_1, \mathbf{n})$. By construction, the z-axis will end up pointing along **n**. As for the x-axis, it must remain perpendicular to the z-axis. If θ_1 varies uniformly over $[0, 2\pi)$, then the image of the x-axis will be a point that varies uniformly over a unit circle that is in the plane containing the origin, and perpendicular to the image of the z-axis. This is easy to see from eq. (41).

Let $Q(\mathbf{v}, \theta_2)$ be the frame corresponding to the FAA parameters **v** and θ_2. We want to consider the mapping from $R(\theta_1, \mathbf{n})$ to $Q(\mathbf{v}, \theta_2)$, where they both represent the same frame (rotation). By construction, we can see that both mappings move the z-axis to the corresponding input vector. Identical rotations move the z-axis to the same location, we must have $\mathbf{v} = \mathbf{n}$. From equation (38), we can conclude that $\theta_2 = \theta_1 + \theta_0(\mathbf{n})$, where $\theta_0(\mathbf{n})$ is a constant dependent on **n**. One can see that $\theta_0(\mathbf{n})$ depends on both the specific details of $R(\mathbf{n})$, and also the arbitrary choices of $\theta = 0$ for the FAA representation (see Definition 6 and its footnote).

As a result, if **n** is uniformly distributed over the unit sphere, **v** will also be uniformly distributed over the unit sphere. Furthermore, eq. (41) tells us that for every **n**, a uniform distribution over θ_1 will induce a uniform distribution over θ_2. Therefore, the rotationally symmetric distribution generated by the subgroup method will induce a uniform distribution over the parameters of the FAA representation, namely $P(\mathbf{v}, \theta)$ as defined above. The FAA representation is an onto and one-to-one mapping from parameters to rotations (see Appendix B). Therefore, the converse is also true: a uniform distribution of FAA parameters, $P(\mathbf{v}, \theta)$, induces a distribution over rotations that is rotationally symmetric. □

LEMMA 2. Given:
1. A rotationally symmetric distribution P over all possible frames.
2. An arbitrary unit vector **v**.

Let U be an arbitrary set of unit vectors (conceptually, this is a patch of the unit sphere).
Let $Q(U)$ = probability that $\mathbf{Hv}$ will fall inside U if we pick a frame **H** according to the distribution P.

We claim that $Q(U)$ is rotationally symmetric, in the sense that $Q(U) = Q(\mathbf{R}U)$ for all **R** where **R** is an arbitrary rotation, and $\mathbf{R}U$ is the set of unit vectors generated by rotating each element of U by **R**.

Proof. We prove this lemma by directly showing that for all sets of unit vectors, U, and for all rotations R, that $Q(U) = Q(\mathbf{R}U)$.

Let H_1 be the set of frames which transform **v** into U, i.e., $H_1\mathbf{v} = U$.
Let H_2 be the set of frames which transform **v** into $\mathbf{R}U$, i.e., $H_2\mathbf{v} = \mathbf{R}U$.

We now show that $\mathbf{R}H_1$ is equal to H_2.

We know that $H_1\mathbf{v} = U$. So, we have $\mathbf{R}H_1\mathbf{v} = \mathbf{R}U$. Therefore, we know that $\mathbf{R}H_1 \subseteq H_2$ because H_2 is defined as the set of all frames which transform **v** into $\mathbf{R}U$.

Similarly, $H_2\mathbf{v} = \mathbf{R}U$ tells us that $\mathbf{R}^{-1}H_2\mathbf{v} = U$. So we know that $\mathbf{R}^{-1}H_2 \subseteq H_1$ because H_1 is defined as the set of all frames which transform **v** into U. Rotations are a one-to-one and onto function from frames to frames, so we can conclude that $H_2 \subseteq \mathbf{R}H_1$.

Since $\mathbf{R}H_1 \subseteq H_2$ and $H_2 \subseteq \mathbf{R}H_1$, we know that $\mathbf{R}H_1 = H_2$.

By rotational symmetry, $P(H_1) = P(\mathbf{R}H_1)$, which in turn means that $P(H_1) = P(H_2)$. Since H_1 maps **v** into U, we have $P(H_1) = Q(U)$. Similarly, $P(H_2) = Q(\mathbf{R}U)$. Together, these imply that $Q(U) = Q(\mathbf{R}U)$. Therefore, Q is rotationally symmetric. The only rotationally symmetric distribution over unit vectors is the uniform distribution over the unit sphere. □

THEOREM 2. The FAA parameterization of frames is isotropically uniform.

Proof. From Lemma 1, we know that a uniform distribution over the parameters of the FAA representation induces a probability distribution over frames which is rotationally symmetric. Thus, we satisfy the second condition of Definition 9.

From Lemma 2, we know that all rotationally symmetric distributions of rotations will randomize any unit vector over the unit sphere in a uniform manner.

Together, we can conclude that a uniform distribution over the parameters of the FAA representation will randomize any unit vector so that it is uniformly distributed over the sphere. Thus, we satisfy the first condition of Definition 9. □

Appendix B: Technical Notes on the FAA Representation

Here we prove two theorems. The first is that the FAA representation is a bijection between $S^2 \times S^1$ and frames. Because $S^2 \times S^1$ and $SO(3)$ have different homology types, they cannot be homeomorphic (Munkres 1984). As a result, any mapping between them cannot simultaneously satisfy all of the following conditions (definition of homeomorphic):

(1) the mapping is continuous;

(2) the mapping's inverse is continuous;

(3) the mapping is one-to-one;

(4) the mapping is onto.

In our case, we prove 3 and 4. So it must be that we cannot satisfy 1 and/or 2. In the second theorem, we give a direct proof that the FAA representation is not continuous.

THEOREM 3. The FAA representation is a one-to-one and onto mapping between $S^2 \times S^1$ and frames.

Proof. To prove that the FAA representation is onto, we show that given any frame $\mathbf{H} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$, there is a unique pair $(\mathbf{v}, \theta)$ which represents that frame. First, observe that $\mathbf{v}$ is always along the z-axis, so $\mathbf{v} = \mathbf{z}$ is uniquely determined. Next, note that $\mathbf{x}$ is a unit vector which is perpendicular to $\mathbf{z}$. Therefore, $\mathbf{x}$ lies on a unit circle perpendicular to $\mathbf{z}$. This, in turn, uniquely specifies the angle θ which represents the direction of the x-axis.

Similarly, given an axis and an angle $(\mathbf{v}, \theta)$, the corresponding frame is uniquely determined. Therefore the mapping is one-to-one. □

Next, we show that the FAA representation is not continuous.

THEOREM 4. The FAA representation is not continuous.

Proof. Our proof will be by contradiction. We will show that if the axis–angle representation were continuous, then it is possible to "continuously comb a sphere with tangent hairs". Since there is a theorem from differential topology (Rotman 1988; Kinsey 1991) showing that this is impossible,[4] it follows that FAA is not a continuous representation of orientations.

Suppose we are given an FAA representation $\mathbf{H}(\mathbf{v}, \theta) = (\mathbf{x}(\mathbf{v}, \theta), \mathbf{y}(\mathbf{v}, \theta), \mathbf{z}(\mathbf{v}, \theta))$. As noted in a footnote earlier, there are an infinite number of FAA representations of orientations, which differ only in their choice of $\theta = 0$ for each $\mathbf{v}$. From $\mathbf{H}$, we construct a set of tangent vectors on the unit sphere S^2. Consider the function $\mathbf{h}(\mathbf{v}) = \mathbf{x}(\mathbf{v}, 0)$. Our function $\mathbf{h}$ is a restricted version of our full function $\mathbf{H}$. So if $\mathbf{h}$ is not continuous, then $\mathbf{H}$ is not continuous.

Our function $\mathbf{h}(\mathbf{v})$ is a mapping from S^2 to unit vectors. We know that $\mathbf{v} = \mathbf{z}(\mathbf{v}, \theta)$ by definition of the FAA representation. We also know that $\mathbf{x}(\mathbf{v}, \theta)$ is perpendicular to $\mathbf{z}(\mathbf{v}, \theta)$ because $\mathbf{H}$ is a mapping to valid frames. Therefore, $\mathbf{h}(\mathbf{v})$ is perpendicular to $\mathbf{v}$, which in turn means $\mathbf{h}(\mathbf{v})$ is tangent to the sphere at $\mathbf{v}$.

So our function $\mathbf{h}$ specifies a complete set of "tangent hairs" on the unit sphere. Therefore, by the above-mentioned theorem from topology, we conclude that $\mathbf{h}$ cannot be continuous. Therefore, $\mathbf{H}$ cannot be continuous. So any FAA representation of orientations must be discontinuous. □

One might wonder if the discontinuities of the FAA representation are a matter of concern. For example, the discontinuities may be localized to some parts of $S^2 \times S^1$, thus making the discontinuities themselves non-isotropic. While this may be true, it is not relevant to our discussion. All we care about is that a uniform distribution over the parameters of FAA induces a distribution of orientations that is isotropically uniform.

A simple analogy would be a parameterization of the unit circle by the interval $[0, 1]$ which is uniform, but not continuous. For example, $t \in [0, 1/2]$ becomes mapped to $2\pi t$ and $t \in [1/2, 1]$ becomes mapped to $2\pi((3/2) - t)$. Although this mapping is not continuous, a uniform distribution over $t \in [0, 1]$ induces a uniform distribution over the unit circle.

Appendix C: Relationships Between Representations of Rotations

We consider the relationship between several different representations of rotations, $SO(3)$. Specifically, we shall look at the representations orthogonal image, frames, quaternions, axis–angle, and the FAA representation. The relationships are summarized by eq. (52).

We begin by showing that $SO(3)$, orthogonal image, and frames are isomorphic to each other. First, consider $SO(3)$ and frames. According to Definition 4, a frame is specified by an ordered triple of unit vectors $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ such that $\mathbf{x} \times \mathbf{y} = \mathbf{z}$. We note that one can convert between $SO(3)$ and frames trivially; given a rotation matrix R in $SO(3)$, the columns of R are the unit vectors of the frame that corresponds to

[4] This result from topology is colloquially known as the Hairy Ball Theorem and is a direct consequence of the famous Brouwer Fixed Point Theorem. See Rotman (1988) and Kinsey (1991).

R. Similarly, given a frame, F, the unit vectors of F are the columns of the corresponding rotation matrix in $SO(3)$. For frames to be isomorphic to $SO(3)$, we need to define the group operator. Naturally, we define multiplication of frames to be the equivalent operation of matrix multiplication on the corresponding matrices. In this way, we have a one-to-one and onto mapping between frames and $SO(3)$ which preserves the group operations on both sides.

Next, we consider the orthogonal image representation of rotations (Mandell et al. 2001; Mitchell 2004). Let R be a rotation, and let $\mathbf{y}$ and $\mathbf{z}$ be the y- and z-axis, respectively. The orthogonal image representation is based on the fact that R is fully specified by the image of the y- and z-axis, namely $R\mathbf{y}$ and $R\mathbf{z}$. The orthogonal image representation is an ordered pair of vectors, which represent the image of the y- and z-axis under the rotation.

If we work in the coordinate frame of $\mathbf{y}$ and $\mathbf{z}$, then $\mathbf{y} = [0\ 1\ 0]^t$ and $\mathbf{z} = [0\ 0\ 1]^t$. Let M be the matrix in $SO(3)$ which represents our rotation R. We now see that $R\mathbf{y}$ and $R\mathbf{z}$ are the second and third columns of M. In fact, the orthogonal image is simply a more compact representation of frames. Frames have a redundant amount of information; given any two vectors of a frame, the third can be uniquely determined by the right-hand rule $\mathbf{x} \times \mathbf{y} = \mathbf{z}$. The orthogonal image representation is similar to frames, except with one vector removed. So if we define the group operator on orthogonal images in the analogous fashion, we see that the orthogonal image representation is isomorphic to frames and thus $SO(3)$.

Now that we have shown $SO(3)$, frames, and the orthogonal image representation to be isomorphic to each other, we consider quaternions. We begin with the mapping from quaternions to $SO(3)$ (Salamin 1979).

$$\text{Let } q = (q_0, q_1, q_2, q_3) \text{ be a quaternion with scalar part } q_0 \text{ and vector part } q_1, q_2, q_3 \tag{42}$$

$$\text{Let } R(q) = \text{the matrix in } SO(3) \text{ corresponding to quaternion } q. \tag{43}$$

$$R(q) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(-q_0q_3 + q_1q_2) & 2(q_0q_2 + q_1q_3) \\ 2(q_0q_3 + q_2q_1) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(-q_0q_1 + q_2q_3) \\ 2(-q_0q_2 + q_3q_1) & 2(q_0q_1 + q_3q_2) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}. \tag{44}$$

As one can see, the mapping from quaterions to $SO(3)$ is continuous in the sense that each component of R is a continuous function of q. One can also observe that q and $-q$ map to the same matrix in $SO(3)$. It is well known that quaternions are a two-to-one homomorphism of $SO(3)$ (Salamin 1979).

Another common representation of rotations is the conventional (classical/canonical) axis–angle, where one specifies the axis of the rotation by a unit vector, and the angle is the amount of rotation around the axis (see Definition 5). The mapping from axis angle to quaternions is

$$\text{Let } T(\mathbf{n}, \theta) = \text{axis–angle representation of a rotation with axis } \mathbf{n} \text{ and angle } \theta. \tag{45}$$

$$\text{Let } q(\theta, \mathbf{n}) = \text{quaternion corresponding to } T(\mathbf{n}, \theta). \tag{46}$$

$$q(\theta, \mathbf{n}) = \cos(\theta/2) + \sin(\theta/2)\mathbf{n}. \tag{47}$$

Notice that the mapping from axis–angle to quaternions is continuous. Furthermore, if we restrict θ to be in the range $[0, 2\pi)$, then the mapping is almost one-to-one. The mapping is one-to-one except when $\theta = 0$.

THEOREM 5. $q(\theta, \mathbf{n})$ is a one-to-one mapping for $\theta \in (0, 2\pi)$, and it is many-to-one when $\theta = 0$.

Proof.

CASE 1: $\theta = 0$.
When $\theta = 0$, the mapping is many-to-one because for any unit vector $\mathbf{n}$, we will map $(\mathbf{n}, 0)$ to the quaternion $[1, 0, 0, 0]$.

CASE 2:$\theta \in (0, 2\pi)$.
Let $(\theta_1, \mathbf{n}_1)$ and $(\theta_2, \mathbf{n}_2)$ be two axis–angle representations of two rotations. To show one-to-one, we need

$$q(\theta_1, \mathbf{n}_1) = q(\theta_2, \mathbf{n}_2) \iff \theta_1 = \theta_2 \text{ and } \mathbf{n}_1 = \mathbf{n}_2. \tag{48}$$

The $\Leftarrow$ direction is trivial.
We show the $\Rightarrow$ direction. Assume that $q(\theta_1, \mathbf{n}_1) = q(\theta_2, \mathbf{n}_2)$.

$$\cos(\theta_1/2) + \sin(\theta/2)\mathbf{n_1} = \cos(\theta_2/2) + \sin(\theta/2)\mathbf{n_2} \tag{49}$$

$$\cos(\theta_1/2) = \cos(\theta_2/2) \tag{50}$$

$$\sin(\theta_1/2)\mathbf{n_1} = \sin(\theta_2/2)\mathbf{n_2}. \tag{51}$$

From eq. (50) and the fact that $\theta_1, \theta_2 \in (0, 2\pi)$, we know $\theta_1 = \theta_2$ because $\cos(\theta_1/2)$ is invertible in this range. From eq. (51) and the fact that $\theta_1 = \theta_2 \neq 0$, we know that $\sin(\theta_1/2) = \sin(\theta_2/2) \neq 0$. So we must have $\mathbf{n}_1 = \mathbf{n}_2$. □

DEFINITION 13. We say a mapping is *nearly one-to-one* if and only if it is one-to-one everywhere, except for a set of measure zero in its range and also for a set of measure zero in its domain.

DEFINITION 14. We say a mapping is *nearly two-to-one* if and only if it is two-to-one everywhere, except for a set of measure zero in its range and also for a set of measure zero in its domain.

So, the angle-axis representation is nearly one-to-one with quaternions, and nearly two-to-one with $SO(3)$. It is well known that for the axis–angle representation of rotations, $(\theta, \mathbf{n})$ and $(2\pi - \theta, -\mathbf{n})$ represent the same rotation.

Finally, we consider the FAA representation. We refer the reader to Appendix B to show that the mapping from the FAA representation to $SO(3)$ is one-to-one, but discontinuous. We summarize the relationships in eq. (52).

$$\begin{tikzcd}
\text{Quaternions} \arrow[r, "{2-1, homomorphism}", "continuous"'] & SO(3) \arrow[d, equal, "isomorphic"] \\
\text{Axis-Angle Rotations} \arrow[u, "nearly\ 1-1", "continous"'] \arrow[r, "nearly\ 2-1", "continuous"'] & \text{Orthogonal Image} \arrow[d, equal, "isomorphic"] \\
\text{Axis-Angle Frames} \arrow[r, "{1-1, bijection}", "discontinuous"'] & \text{Frames}
\end{tikzcd} \tag{52}$$

We close by noting a few of the representations which are isotropically uniform. For quaternions, a uniform parameterization of the three sphere S^3 will sample rotations in a manner which is isotropically uniform (Salamin 1972). As shown in this paper, the FAA representation is isotropically uniform. However, the conventional (canonical) axis–angle representation of rotations is not isotropically uniform (see the end of Appendix A). The orthogonal image method is also known to be isotropically uniform (Mandell et al. 2001; Mitchell 2004) (see equation 52).

Acknowledgments

The authors are especially grateful to Dr J. C. Mitchell, who provided mathematical advice and critical assistance on our manuscript. This work is supported by grants to BRD from the National Institutes of Health (R01 GM-65982), and the National Science Foundation (IIS-9906790, EIA-0102710, EIA-0102712, EIA-9818299, EIA-9802068, and EIA-0305444).

References

Al-Hashimi, H. M., and Patel, D. J. 2002. Residual dipolar couplings: synergy between NMR and structural genomics. *Journal of Biomolecular NMR* 22(1):18.

Al-Hashimi, H. M., Gorin, A., Majumdar, A., Gosser, Y., and Patel, D. J. 2002. Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. *Journal of Molecular Biology* 318:637–649.

Andrec, M., Du, P., and Levy, R. M. 2001. Protein backbone structure determination using only residual dipolar couplings from one ordering medium. *Journal of Biomolecular NMR* 21:335–347.

Artymiuk, P. J., Blake, C. C. F., Rice, D. W., and Wilson, K. S. 1982. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 angstroms resolution. *Acta Crystallographica B* 38:778.

Babu, C. R., Flynn, P. F., and Wand, A. J. 2001. Validation of protein structure from preparations of encapsulated proteins dissolved in low viscosity fluids. *Journal of the American Chemical Society* 123:2691.

Bailey-Kellogg, C., Widge, A., Kelley, J. J. III, Berardi, M. J., Bushweller, J. H., and Donald, B. R. 2000. The NOESY jigsaw: automated protein secondary structure and mainchain assignment from sparse, unassigned NMR data. *Journal of Computational Biology* 7(3–4):537–558.

Chen, Y., Reizer, J., Saier M. H. Jr, Fairbrother, W. J., and Wright, P. E. 1993. Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc. *Biochemistry* 32(1):32–37.

Clore, G. M., Gronenborn, A. M., and Bax, A. 1998. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *Journal of Magnetic Resonance* 133:216–221.

Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society* 120:6836–6837.

Delaglio, F., Kontaxis, G., and Bax, A. 2000. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *Journal of the American Chemical Society* 122:2142–2143.

Diaconis, P., and Shahshahani, M. 1999. The subgroup algorithm for generating uniform random variables. *Problems in Engineering and Information Science* 1:15–32.

Diamond, R. 1974. Real-space refinement of the structure of hen egg-white lysozyme. *Journal of Molecular Biology* 82:371–391.

Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. 2002. NMR analysis of a 900K GroEL–GroES complex. *Nature* 418:207–211.

Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay, Murcko, M. A., and Moore, J. M. 1999. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chemistry and Biology* 6:755–769.

Fowler, C. A., Tian, F., Al-Hashimi, H. M., and Prestegard, J. H. 2000. Rapid determination of protein folds using residual dipolar couplings. *Journal of Molecular Biology* 304(3):447–460.

Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G. L. 1994. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* 33:4721–4729.

Giesen, A. W., Homans, S. W., and Brown, J. M. 2003. Determination of protein global folds using backbone residual dipolar coupling and long-range NOE restraints. *Journal of Biomolecular NMR* 25(1):63–71.

Girard, E., Chantalat, L., Vicat, J., and Kahn, R. 2001. Gd-HPDO3A, a complex to obtain high-phasing-power heavy atom derivatives for SAD and MAD experiments: results with tetragonal hen egg-white lysozyme. *Acta Crystallographica D* 58:1–9.

Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* 253:657.

Hus, J. C., Propmers, J., and Brüschweiler, R. 2002. Assignment strategy for proteins of known structure. *Journal of Magnetic Resonance* 157:119–125.

Huttenlocher, D. P., and Kedem, K. 1992. Distance metrics for comparing shapes in the plane. *Symbolic and Numerical Computation for Artificial Intelligence*, Chapter 8, B. R. Donald, D. Kapur, and J. Mundy, editors. Academic Press, Harcourt Jovanovich, London, pp. 201–219.

Huttenlocher, D. P., and Kedem, K. 1990. Computing the minimum Hausdorff distance for point sets under translation. *Proceedings of the 6th ACM Symposium on Computational Geometry*, Berkeley, CA, pp. 340–349.

Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. 1999. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* 7(8):967–976.

Kendall, M. G., and Moran, P. A. P. 1963 *Geometrical Probability*. Griffin, London, p. 93.

Kinsey, L. C. 1991. Miscellany. *Topology of Surfaces*, chapter 10. Springer-Verlag, Berlin, pp. 199–203.

Kurinov, I. V., and Harrison, R. W. 1995. The influence of temperature on lysozyme crystals – structure and dynamics of protein and water. *Acta Crystallographica D* 51:98–109.

Kuszewski, J., Gronenborn, A. M., and Clore, G. M. 1999. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *Journal of the American Chemistry Society* 121:2337–2338.

Langmead, C. J., and Donald, B. R. 2004. An expectation/maximization nuclear vector replacement algorithm for automated NMR resonance assignments. *Journal of Biomolecular NMR* 29(2):111–138.

Langmead, C. J., Yan, A. K., Wang, L., Lilien, R., and Donald, B. R. 2003. A polynomial time nuclear vector replacement algorithm for automated NMR resonance assignments. *Proceedings of the 7th Annual International Conference on Computational Molecular Biology (RECOMB)*, Berlin, Germany, April 10–13, pp. 176–187.

Langmead, C. J., Yan, A. K., Wang, L., Lilien, R., and Donald, B. R. 2004. A polynomial time nuclear vector replacement algorithm for automated NMR resonance assignments. *Journal of Computational Biology* 11(2–3):277–298.

Lim, K., Nadarajah, A., Forsythe, E. L., and Pusey, M. L. 1998. Locations of bromide ions in tetragonal lysozyme crystals. *Acta Crystallographica D* 54:899–904.

Losonczi, J. A., Andrec, M., Fischer, W. F., and Prestegard, J. H. 1999. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance* 138(2):334–342.

Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyi, V., Mitchell, J. C., Nelson, E., Tsigelny, I., and Eyck, L. F. 2001. Protein docking using continuum electrostatics and geometric fit. *Protein Engineering* 14:105.

Mitchell, J. C. 2004. The orthogonal image method for generating uniform distributions of rotation matrices. *Proteins: Structure, Function and Genetics* submitted for publication.

Munkres, J. R. 1984. *Elements of Algebraic Topology*. Addison-Wesley, Reading, MA.

Oki, H., Matsuura, Y., Komatsu, H., and Chernov, A. A. 1999. Refined structure of orthorhombic lysozyme crystallized at high temperature: correlation between morphology and intermolecular contacts. *Acta Crystallographica D* 55:114.

Palmer, A. G. III. 1997. Probing molecular motion by NMR. *Current Opinion in Structural Biology* 7:732–737.

Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S., and Shaw, K. 1994. Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Journal of Biochemistry* 299:151–158.

Rohl, C. A., and Baker, D. 2002. De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *Journal of the American Chemistry Society* 124:2723–2729.

Rossman, M. G., and Blow, D. M. 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15:24–31.

Rotman, J. J. 1988. Excision and applications. *An Introduction to Algebraic Topology*, chapter 6. Springer-Verlag, Berlin, p. 123.

Salamin, E. 1972. Quaternions. HAKMEM. Artificial Intelligence Memo No. 239, Massachusetts Institute of Technology, AI Laboratory, M. Beeler et al., editors, February 29. Item 107. http://www.inwap.com/pdp10/hbaker/hakmem/hakmem.html.

Salamin, E. 1979. Applications of quaternions to computation with rotations. Internal working paper, Stanford University, AI Laboratory.

Saupe, A. 1968. Recent results in the field of liquid crystals. *Angewandte Chemie* 7:97–112.

Schwalbe, H., Grimshaw, S. B., Spencer, A., Buck, M., Boyd, J., Dobson, C. M., Redfield, C., and Smith, L. J. 2001. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Science* 10:677–688.

Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. 1996. Discovering high affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534.

Tian, F., Valafar, H., and Prestegard, J. H. 2001. A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *Journal of the American Chemistry Society* 123:11,791–11,796.

Tjandra, N., and Bax, A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114.

Vaney, M. C., Maignan, S., Ries-Kautt, M., and Ducruix, A. 1996. High-resolution structure (1.33 angstrom) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. *Acta Crystallographica D* 52:505–517.

Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. 1987. Structure of ubiquitin refined at 1.8 A resolution. *Journal of Molecular Biology* 194:531–544.

Wang, L., and Donald, B. R. 2004. Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure. *Journal of Biomolecular NMR* 29(3):223–242.

Wedemeyer, W. J., Rohl, C. A., and Scheraga, H. A. 2002. Exact solutions for chemical bond orientations from residual dipolar couplings. *Journal of Biomolecular NMR* 22:137–151.

Zimmerman, D. E., Kulikowski, C. A., Feng, W., Tashiro, M., Chien, C-Y., Ros, C. B., Moy, F. J., Powers, R., Montelione, G. T. 1997. Artificial intelligence methods for automated analysis of protein resonance assignments. *Journal of Molecular Biology* 269:592–610.