Published in final edited form as:

J Phys Chem B. 2009 May 7; 113(18): 6173–6176. doi:10.1021/jp900411z.

Approximate Reconstruction of Continuous Spatially Complex Domain Motions by Multi-alignment NMR Residual Dipolar Couplings

Charles K. Fisher and Hashim M. Al-Hashimi*

Department of Chemistry & Biophysics The University of Michigan, 930 North University Avenue
Ann Arbor, MI 48109-1055, USA Tel: 734-615-3361, Fax: 734-647-4865

Abstract

NMR spectroscopy is one of the most powerful techniques for studying the internal dynamics of biomolecules. Current formalisms approximate the dynamics using simple continuous motional models or models involving discrete jumps between a small number of states. However, no approach currently exists for interpreting NMR data in terms of continuous spatially complex motional paths that may feature more than one distinct maneuver. Here we present an approach for approximately reconstructing spatially complex continuous motions of chiral domains using NMR anisotropic interactions. The key is to express Wigner matrix elements, which can be determined experimentally using residual dipolar couplings, as a line integral over a curve in configuration space containing an ensemble of conformations and to approximate the curve using a series of geodesic segments. Using this approach and five sets of synthetic residual dipolar couplings computed for five linearly independent alignment conditions, we show that it is theoretically possible to reconstruct salient features of a multi-segment inter-helical motional trajectory obtained from a 65 ns molecular dynamics simulation of a stem-loop RNA. Our study shows that the 3-D atomic reconstruction of complex motions in biomolecules is within experimental reach.

Keywords

dynamics; Wigner; trajectory; internal motions; RDC; quaternion

An outstanding challenge in biophysics is the development of experimental methods for visualizing the internal dynamics of biomolecules at atomic resolution. NMR spectroscopy is one of few techniques that can be used to probe internal motions throughout a molecule with site-specific resolution¹⁻³. Currently, the internal dynamics of biomolecules cannot be reconstructed based on NMR data alone. Rather, NMR data is typically interpreted using simplified motional models that approximate the real dynamics⁴. This includes Gaussian fluctuations, motions within a cone, and discrete jumps between a small number of states⁵⁻⁸. No approach currently exists for interpreting NMR data in terms of more spatially complex continuous motional trajectories that may encompass not one, but a series of distinct maneuvers or segments. Obtaining insights into these spatial details can be essential for understanding how motions contribute to folding and function. For example, for many enzymes, internal motions of groups lining the active sites are believed to be spatially optimized to allow substrate

*Email: hashimi@umich.edu.

Supporting Information Available. The MD and reconstructed trajectory as a function of the inter-helical bend angles and the effect of experimental uncertainty are illustrated in Figures S1 and S2, respectively. A comparison of the MD trajectory and the approximate reconstruction is illustrated in Movie S1. This information is available free of charge via the Internet at <http://pubs.acs.org>.

entry and product release^{9,10} and formation of catalytically competent states¹¹. Helices in RNA have also been shown to undergo motions that are spatially tuned to allow recognition of diverse targets¹². Likewise, domains in multi-domain proteins are believed to undergo rigid-body motions that are spatially optimized to carry out diverse transactions^{13,14}.

The challenge in visualizing dynamics by NMR is that the number of parameters needed to fully characterize a complex motional path typically far exceeds the number of parameters that can be measured experimentally. We recently showed¹⁵ that a maximum of twenty-five motionally averaged Wigner rotation elements can theoretically be determined for chiral domains from measurements of residual anisotropic interactions such as residual dipolar couplings (RDCs)^{7,16} under five linearly independent alignment conditions. The twenty-five elements significantly expand the parameters available for describing motional paths, which previously were largely limited to five Wigner matrix elements per chiral domain¹⁷ or bond vector^{18,19}. These elements define the spatial resolution limit with which domain motions can be characterized using RDCs^{15,20}. In this letter, we describe a formalism that allows one to fully exploit all twenty-five elements in the atomic reconstruction of complex continuous motional trajectories. We focus on motions of chiral fragments such as domains in proteins and helices in RNA that can trace out spatially complex motional trajectories.

The second rank Wigner rotation elements, D_{mk}^2 , are commonly parameterized as a function of three Euler angles ($\alpha\beta\gamma$) that specify the orientation of a domain relative to a reference frame.

In the presence of motions, the time-averaged Wigner elements, $\langle D_{mk}^2(\alpha\beta\gamma) \rangle$, represent population-weighted averages over all orientations sampled. Here, we utilize the quaternion (q) representation of the Wigner matrix elements, $\langle D_{mk}^2(q) \rangle$, for its mathematical simplicity^{15,21,22}. The latter are 4-dimensional complex numbers (q_0, q_1, q_2 , and q_3) that represent an orientation as a single axis rotation^{21,23}; q_0 encodes the rotational amplitude and q_1, q_2 , and q_3 encode the x, y and z components of the rotation axis, respectively. The set of unit quaternions defines the surface of a 4-dimensional hyper-sphere, referred to as S^3 , with a 2-1 mapping to the rotation group²³. Thus, the distribution of domain orientations can be described as a distribution on S^3 .

Our approach is to model the distribution of domain orientations as a continuous path on S^3 . The path is parameterized using a function, $q(u)$, which is a unit quaternion specifying the domain orientation as a function of a trajectory coordinate, u . We assume that each orientation along the path is equally weighted. Thus, the resultant net population-weight (P_j) for a given orientation will be proportional to the number of times (n_j) the path visits that particular orientation. There may be distinct paths linking various domain orientations in the ensemble such that $n_j \propto P_j$; we will refer to this as a degeneracy of the first kind (d_1). In addition, there may be paths that produce the same motionally averaged Wigner elements despite representing different orientational distributions; we refer to this as a degeneracy of the second kind (d_2).

The average value of the Wigner matrix elements over a distribution of orientations can be calculated as a line integral over the path on S^3 representing that distribution, provided that $q(u)$ is a piecewise differentiable function, as

$$\langle D_{mk}^2(q) \rangle = \frac{1}{\int_a^b \|q'(u)\| du} \int_a^b D_{mk}^2(q(u)) \|q'(u)\| du \quad (1)$$

where $q'(u)$ is the derivative of $q(u)$, the double brackets denote a vector magnitude, and a and b denote the starting and final position along the path, respectively. In words, eq. 1 yields the

sum over the values of $D_{mk}^2(q)$, weighted by the arc length, at all points along the path, divided by the total length of the path.

In the analysis of domain motions by NMR we aim to find a function, $q(u)$, that satisfies eq. 1 for the experimentally determined Wigner matrix. In order to make the search simpler we choose $q(u)$ to be of a particular functional form that is general enough to describe arbitrarily complex motions yet is defined by a relatively small number of parameters. Typically, this choice of $q(u)$ will be an approximation of the actual distribution of orientations. First, we express $q(u)$ as a union of a finite number of n differentiable segments each defined by $q_i(u_i)$ (Figure 1),

$$\langle D_{mk}^2(q) \rangle \approx \frac{1}{\sum_{i=1}^n \int_{a_i}^{b_i} \|q'_i(u_i)\| du} \sum_{i=1}^n \int_{a_i}^{b_i} D_{mk}^2(q_i(u_i)) \|q'_i(u_i)\| du_i. \quad (2)$$

Next, we choose segments corresponding to the shortest path, or geodesic, between the start and end points, $q_{i,a}$ and $q_{i,b}$ respectively,

$$q_i(u_i) = \frac{\sin((1-u_i)\omega_i)}{\sin(\omega_i)} q_{i,a} + \frac{\sin(u_i\omega_i)}{\sin(\omega_i)} q_{i,b}, \quad (3)$$

where $\omega_i = \arccos(q_{i,a} \cdot q_{i,b})$ is the length of the i^{th} segment, and u_i is its trajectory coordinate that ranges from zero to one²³. Substituting eq. 3 into eq. 2 yields

$$\langle D_{mk}^2(q) \rangle \approx \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \int_0^1 D_{mk}^2(q_i(u_i)) \omega_i du_i, \quad (4)$$

where the endpoints of the segment are chosen such that $q_{i+1,a} = q_{i,b}$ to ensure continuity (Figure 1).

By finding the endpoints along each segment that back-predict the experimentally determined average Wigner elements, we can directly solve for a segmented path approximately describing motions of the domain based on multi-alignment RDCs using eq. 4. The number of segments that can be used is limited by the number of Wigner elements that can be measured. Each quaternion is specified by three parameters. Thus, a model consisting of n segments requires $3n + 3$ parameters. A maximum of twenty-five Wigner elements can be measured from RDCs such that models consisting of more than seven segments are inherently underdetermined by the RDCs. Fitting models with an increasing number of parameters rapidly becomes computationally expensive due to the increase in the number of expressions that need to be evaluated and presence of many local minima.

While there is no information regarding the temporal ordering of states along the path, we can choose to order the states assuming a unidirectional trajectory starting from $q_{1,a}$ and ending at $q_{n,b}$. This “pseudo-trajectory” (or its reverse), illustrates a probable path but ignores the stochastic nature of the fluctuations as they occur in solution. The time-dependence can thus be modeled by starting the trajectory at time $t = 0$ at position $q_{1,a}$ (or $q_{n,b}$) and ending the trajectory at time $t = \tau_f$ at position $q_{n,b}$ (or $q_{1,a}$). The fraction of time spent in the i^{th} segment (p_i) is proportional to the length of the i^{th} segment (ω_i) divided by the total length ($\sum_{j=1}^n \omega_j$).

Defining $\rho_i = \omega_i / \sum_{j=1}^n \omega_j$ and $\rho_0 = 0$ we can make a substitution for u_i in terms of t as

$u_i = \frac{t - \sum_{j=0}^{i-1} \rho_j \tau_f}{\rho_i \tau_f}$; $\sum_{j=0}^{i-1} \rho_j \tau_f \leq t \leq \sum_{j=0}^i \rho_j \tau_f$ which moves the conformation from the beginning to the end of the path according to the model criteria.

We devised a successive fitting procedure utilizing adaptive nonlinear optimization methods, such as the simulated annealing (SA) or genetic algorithms (GA), to minimize a reduced chi-squared statistic, χ^2 , comparing the theoretical and experimental Wigner matrix elements. The reduced chi-squared statistic was calculated as,

$$\chi^2 = \sum_{m=-2}^2 \sum_{m=0-1; m \leq 0} \left(\frac{\left(\text{Re} \left(\langle D_{mm}^2(q) \rangle_{fit} \right) - \text{Re} \left(\langle D_{mm}^2(q) \rangle_{exp} \right) \right)^2 + \left(\text{Im} \left(\langle D_{mm}^2(q) \rangle_{fit} \right) - \text{Im} \left(\langle D_{mm}^2(q) \rangle_{exp} \right) \right)^2}{\sigma^2(p=v)} \right)$$

where p is the number of “experimental” Wigner elements, n is the number of degrees of freedom in the model, and σ^2 is the standard deviation in the measurements. Our approach, outlined below, aims to find the simplest model that reproduces the data within the experimental error by assuming that the addition of a new segment to the approximation results in an incremental improvement in the path reconstruction.

1. Find the first segment defined by the two quaternions, $q_{1,a}$ and $q_{1,b}$, that minimize χ^2 using an SA or GA algorithm.
2. If $\chi^2 < 1$, stop. If $\chi^2 \geq 1$, continue to step 3.
3. For the next approximation, successively increase the number of segments n defined by quaternions $q_{i,a}$ and $q_{i,b}$, for $i \in \{1, 2, \dots, n\}$ to find that path that minimize χ^2 . Here, initial values for $q_{i,a}$ and $q_{i,b}$ are chosen to roughly uniformly sample the path obtained from the prior $n-1$ segment approximation.
4. If the degrees of freedom are exhausted continue to step 5; otherwise go back to step 2.
5. If $\chi^2 = 0$, stop. If $\chi^2 > 0$, no model will fit the data within the designated error. Simulations of simple motional paths of geodesic segments were used to validate equations and the procedure described above (data not shown).

As an illustration of our method, we reconstructed rigid-body motions of two helices obtained from a 65 ns molecular dynamics (MD) simulation of HIV-1 transactivation response element (TAR) RNA^{15,24}. TAR is an RNA stem-loop consisting of two helices that are connected by a trinucleotide bulge linker. Inter-helical Euler angles ($\alpha_h \beta_h \gamma_h$) describing the relative orientation of the two helices¹² were computed for each of 64998 TAR snap shots obtained every picosecond from the 65 ns MD trajectory²⁴, as described previously¹⁵. The angles α_h and γ_h specify rotations around the two helical axes and β_h is the inter-helical bend angle. The twenty-five $\langle D_{mk}^2(q) \rangle$ elements describing rigid-body helix motions were then computed by averaging over all MD snap-shots, as described previously¹⁵. The $\langle D_{mk}^2(q) \rangle$ elements, which can be determined experimentally using five linearly independent sets of RDCs¹⁵, were then used to back-predict the approximation of the motional trajectory using eq. 4 and the successive fitting procedure implemented using the computer algebra system Mathematica²⁵.

As shown in Figure 2(a), the χ^2 goodness-of-fit statistic decreases with successive segments until the maximum of seven segments is reached. As shown in Figure 2(b), each point along the reconstructed multi-segment trajectory falls within the distribution of inter-helical orientations sampled during the MD trajectory. The reconstructed path broadly samples the MD conformations. Remarkably, even the time ordering of the states is fairly accurately

reproduced, as shown in Figure 2(c) by a quaternion representation of the time evolution of the inter-helical TAR conformation (also see Movie S1 and Figure S1 in the Supplementary Material). This suggests that the MD trajectory generally follows the equal weight along the path and unidirectionality assumptions inherent to our approach. The d_1 type degeneracy, time-reversal, was the only degeneracy encountered in reconstructing the motions in these simulations, which assume zero experimental uncertainty.

Another way to compare the MD and NMR reconstructed trajectories is to compare the relative populations of different inter-helix orientations. In particular, it is instructive to compare the potential of mean force (F)²⁶ for the two trajectories that provides a measure of the relative populations of different conformations and can be used as an estimate of the change in free energy over a reaction coordinate²⁶. F can be estimated from a trajectory by counting the number of times, n_j , a region of conformational space is sampled along the path. F was calculated as $F(q^i, j) = -kT \ln(n_j) + C^{26}$, where $i \in \{0, 1, 2, 3\}$ specifies the quaternion element, j specifies the bin and C is a constant which is typically chosen such that the minimum of F is zero²⁶. A comparison of the MD and reconstructed F is shown in Figure 3 for each of the quaternion elements. Our approach performs remarkably well in representing the overall conformational distribution and thus the free energy profile (Figure 3); however, it is clear that fine details of the ensemble cannot be captured with only seven segments leading to narrow potentials of mean force and underestimation of the entire range of conformations sampled. In particular, the range with which states can adopt unequal weighting is limited such that lowly populated states may not be adequately captured.

How uniquely can one define motional paths using our approach? In general, d_1 type degeneracies cannot be eliminated because the same distribution can arise from numerous stochastic trajectories; however, since d_1 paths describe the same underlying orientational distribution they are trivial degeneracies. In contrast, d_2 type degeneracies are defined as “incorrect” paths that still reproduce the data. While we did not encounter d_2 in simulations aimed at reconstructing simple paths (data not shown) or the MD trajectory, these degeneracies may arise and become problematic when treating real data containing experimental uncertainty. Though a complete analysis of error propagation is beyond the scope of this work, simulations provided in Supplementary Material suggest that the levels of uncertainty typically encountered in the analysis of RNA A-form helices²⁷ result in fairly small perturbations for $m \leq 3$ (Figure S2). Ideally, all paths that reproduce the experimental measurements within the uncertainty should be considered. Limiting the conformational space by inclusion of loose physical restraints¹² or sampling from MD generated distributions²⁸ may also help overcome d_2 degeneracies.

In conclusion, we described a general approach for analyzing anisotropic NMR interactions in terms of continuous motions of arbitrary form and complexity. The inclusion of other experimental data and integration of computational simulations will help improve the resolution limit for motional characterization beyond that which can be achieved using RDC data alone. Our framework also allows deconvolution of complex motional trajectories derived from MD and other methods into series of successive constitutive maneuvers that can help define the underlying motional choreography.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Al-Hashimi lab for insightful comments and Dr. Ioan Andricioaei (University of California, Irvine), Aaron Frank, and Dr. Catherine Musselman for providing us with the TAR MD trajectory. This work was

supported by a National Science Foundation CAREER award (MCB 0644278) and by funding from the NIH (RO1 AI066975-01) to H.M.A.

References

1. Mittermaier A, Kay LE. *Science* 2006;312:224. [PubMed: 16614210]
2. Palmer AG 3rd, Massi F. *Chem. Rev* 2006;106:1700. [PubMed: 16683750]
3. Getz M, Sun X, Casiano-Negroni A, Zhang Q, Al-Hashimi HM. *Biopolymers* 2007;86:384. [PubMed: 17594140]
4. Bruschweiler R. *Curr. Opin. Struct. Biol* 2003;13:175. [PubMed: 12727510]
5. Bruschweiler R, Wright PE. *J. Am. Chem. Soc* 1994;116:8426.
6. Bernado P, Blackledge M. *J. Am. Chem. Soc* 2004;126:4907. [PubMed: 15080696]
7. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. *Nat. Struct. Bio* 1997;4:292. [PubMed: 9095197]
8. Clore GM, Schwieters CD. *Biochemistry* 2004;43:10678. [PubMed: 15311929]
9. Wang LC, Pang YX, Holder T, Brender JR, Kurochkin AV, Zuiderweg ERP. *Proc. Natl. Acad. Sci. U. S. A* 2001;98:7684. [PubMed: 11438724]
10. Chang CE, Shen T, Trylska J, Tozzini V, McCammon JA. *Biophys. J* 2006;90:3880. [PubMed: 16533835]
11. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D. *Nature* 2007;450:838. [PubMed: 18026086]
12. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. *Nature* 2007;450:1263. [PubMed: 18097416]
13. Bruschweiler R, Liao X, Wright PE. *Science* 1995;268:886. [PubMed: 7754375]
14. Ryabov YE, Fushman D. *J. Am. Chem. Soc* 2007;129:3315. [PubMed: 17319663]
15. Fisher CK, Zhang Q, Stelzer A, Al-Hashimi HMJ. *Phys. Chem. B*. 2008 In Press. (Published online).
16. Tjandra N, Bax A. *Science* 1997;278:1111. [PubMed: 9353189]
17. Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH. *J. Am. Chem. Soc* 2001;123:1416. [PubMed: 11456715]
18. Peti W, Meiler J, Bruschweiler R, Griesinger C. *J. Am. Chem. Soc* 2002;124:5822. [PubMed: 12010057]
19. Briggman KB, Tolman JR. *J. Am. Chem. Soc* 2003;125:10164. [PubMed: 12926926]
20. Zhang Q, Al-Hashimi HM. *Nat. Methods* 2008;5:243. [PubMed: 18246076]
21. Blumich B, Spiess HW. *J. Magn. Reson* 1985;61:356.
22. Emsley L, Bodenhausen G. *J. Magn. Reson* 1992;97:1.
23. Shoemake K. *SIGGRAPH* 1985;19:3.
24. Musselman C, Al-Hashimi HM, Andricioaei I. *Biophys J* 2007;93:411. [PubMed: 17449677]
25. Wolfram Research, Inc.. *Mathematica Edition: Version 6.0*. Wolfram Research, Inc.; Champaign, Illinois: 2007.
26. Leach, Andrew R. *Molecular Modeling: Principles and Applications*. Second Edition. Pearson Education Limited; Harlow, England: 2001.
27. Musselman C, Pitt SW, Gulati K, Foster LL, Andricioaei I, Al-Hashimi HM. *J. Biomol. NMR* 2006;36:235. [PubMed: 17077936]
28. Chen Y, Campbell SL, Dokholyan NV. *Biophys. J* 2007;93:2300. [PubMed: 17557784]

Figure 1.
Illustration of a complex trajectory (yellow) on a 3-dimensional sphere and a segmented approximation (red).

Figure 2.

Reconstruction of inter-helical motions derived from a 65 ns MD simulation of TAR RNA. (a)

Plot of χ^2/χ_{low}^2 versus the number of segments used to reconstruct the inter-helical dynamics.

The χ^2 was normalized relative to the lowest value (seven segments) because there is no measurement error in the MD simulation. (b) Comparison of MD snap-shots (grey spheres) and the pseudo-trajectory that is reconstructed (blue line) based on 25 experimentally

accessible $\langle D_{mk}^2(q) \rangle$ elements. Shown are inter-helical Euler angles (α, β, γ) defining the relative orientation of the helices in each conformation. (c) Comparison of four quaternion elements for the pseudo-(blue line) and MD (grey line) trajectories as a function of time.

Figure 3. Comparison of potentials of mean force calculated from the MD trajectory (in gray) and the best-fit pseudo-trajectory (in blue). A bin size of 0.01 was used for each of the quaternion elements. Lower values of F correspond to a greater population.