J|A|C|S
COMMUNICATIONS

Published on Web 01/28/2004

Periodicity in Residual Dipolar Couplings and Nucleic Acid Structures

Joseph D. Walsh, Javier Cabello-Villegas, and Yun-Xing Wang*

Protein-Nucleic Acid Interaction Section, Structural Biophysics Laboratory, National Cancer Institute, National Institutes of Health, Frederick, Maryland 21702

Received November 5, 2003; E-mail: wangyu@ncifcrf.gov

Analysis of residual dipolar couplings (RDCs) in terms of "RDC waves" allows global alignment and local structure information to be extracted because of the greatly reduced RDC degeneracy problem. RDCs that are measured in weakly aligned samples have been used to validate, refine, and assess three dimensional (3D) topology of protein structure[1,2] and to a lesser extent to refine nucleic acid structures.[3] This is because RDCs contain bond orientation information relative to the external magnetic field. Recent protein studies both in the solid state and in solution demonstrate that the periodicity of static dipolar couplings and RDCs are intrinsic to such structural elements such as α-helices[4,5] and β-sheets.[6] In nucleic acids, one of the main secondary structural elements is the duplex (Figure 1) whose periodicity is reflected in its RDCs, as we show here. In general the RDC, D_{AB}, is expressed in terms of the bond vector orientation in the alignment tensor system. When considering structural elements of known types, such as the duplex in nucleic acids, D_{AB} can be expressed in terms of the bond vector orientation (δ, ρ) in the duplex reference frame and the orientation of the duplex (Θ, Φ) in the alignment frame[7]

$$D_{AB} = C_1 \cos 2\rho_n + C_2 \sin 2\rho_n + C_3 \cos \rho_n + C_4 \sin \rho_n + C_5 \quad (1)$$

where $\rho_n = (\alpha_n + \rho_0)$ is the phase of the bond vector of the nth residue with initial bond vector phase ρ_0 (Figure 1). The slant angle δ_n is the angle the AB bond vector makes with the duplex axis and coefficients $C_i = C_i(\Theta, \Phi, \delta_n)$ (Supporting Information).

While eq 1 is universally applicable to periodic structures such as α-helices and β-strands in peptides, in the case of nucleic acids it is particular intriguing and worthy of investigation. This is because the predominant secondary structure element is the duplex. It exhibits 11 and 10 base pairs per turn for the A- and B-form, respectively, as compared to 3.6 residues per turn in the peptide α-helix. Furthermore, a nucleic acid duplex consists of two helical strands, which are opposite to each other in both the longitudinal polarity and the transverse phase. In the structure determination of nucleic acids using NMR, almost all of the restraints, except those defining the base pairs, are short ranged, such as J couplings and intra-base pair or sequential correlations. In contrast, the structural information extracted from the "dipolar wave" [4,8] of the duplex would allow a better determination of both the global and local conformations. This becomes particularly valuable when considering that both the global and local conformations of nucleic acids are underdetermined with the traditional NMR parameters because nucleic acid structures tend to be elongated and have a very low proton-spin density. The strategy of utilizing the RDCs of nucleic acids in terms of the "RDC wave" has several advantages for the reasons that we demonstrate in this communication.

Since there are more residues per turn in the duplex than in the α-helix in peptides, the RDCs are sampled more often over their range, and hence subtle features of the "RDC wave" can emerge

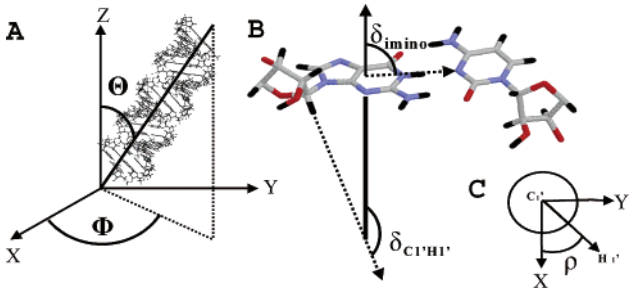

Figure 1. (A) A-form RNA duplex oriented at angles (Θ, Φ) with respect to the magnetic alignment frame (X, Y, Z). (B) G−C base pair tilted for clarity to show the imino and $C_1'H_1'$ slant angles with respect to the helix axis. (C) Looking down the helix axis, the $C_1'H_1'$ angle ρ is shown.

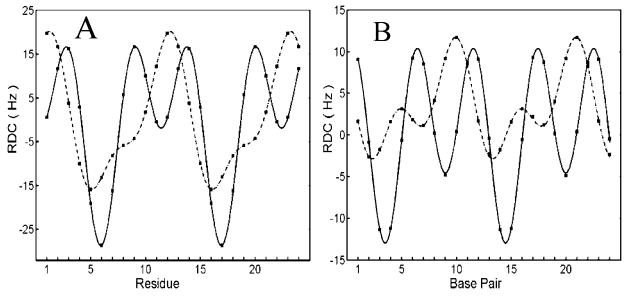

Figure 2. Simulated RDCs from an ideal duplex (Quanta, Accelrys Inc.) A-RNA 24-mer (circles) fit with the helical model of eq 1 with $D_a = -14.5$ Hz, $R = 0.5$. (A) Fits to the $C_1'H_1'$ RDCs for two helix orientations with respect to the principle alignment frame, (Θ, Φ) = (20°, 60°) (Dashed line), and (Θ, Φ) = (60°, 0°) (solid line). (B) Simulated dipolar wave and fits for the imino group with the same duplex orientations as in A. All fits have standard deviation <0.15 Hz.

(see $C_1'H_1'$ and imino "RDC waves", Figure 2, A and B). These features are strongly dependent on the helix orientation angles Θ and Φ and hence are indicative of how the duplex is orientated in the alignment tensor system. Another unique feature of the duplex is that the imino bond vector on one strand points in the exact opposite direction from its orientation were it to be located at the position of its base pair (Figure 1B). Because of this, "RDC waves" are observed for imino groups (Figure 2B and 3A), even though the imino RDC values may originate partly from one strand and partly from the other. Comparison of A and B of Figure 2 shows that the "RDC wave" is qualitatively similar for $C_1'H_1'$ and the iminos, although as expected the phases and amplitudes differ significantly because of the difference in their slant angles and gyromagnetic ratios.

The utility of using eq 1 to fit experimental "RDC wave" data to extract the duplex orientation is evident in the example of the Negative Regulator of Splicing RNA fragment (NRS23). The NRS23 contains less than one full turn of a stem, interrupted by a bulge and a small loop (unpublished results). The imino RDCs of the base-paired region were fit using a fit A-RNA imino slant angle of 97°. The duplex orientation extracted from the fit was (Θ, Φ)

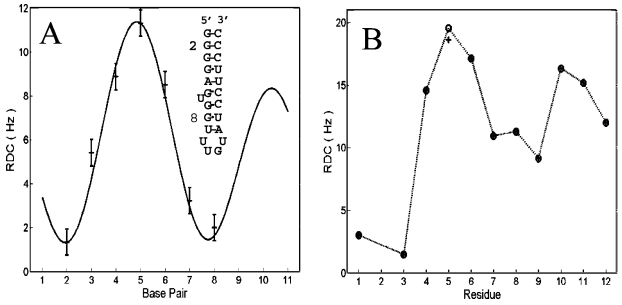

Figure 3. (A) Experimental NRS23 imino RDCs (+) from the base-pairing region showing 1.2 Hz error bars. The duplex fit (line) yielded (Θ, Φ) = (166°, 86°), compared with the values obtained from Saupe matrix elements determined by SVD: (Θ, Φ) = (164°, 73°). (B) Experimental $C_1'H_1'$ RDCs of the Dickerson dodecamer (circles) with fit (Θ, Φ) = (23°, 155°) (+) and fit slant angles having a correlation coefficient of $R = 0.90$ with the experimental slant angles. The size of the markers is greater than the experimental standard deviation of 0.2 Hz[9] and the fitted and experimental data coincide with each other except for residue 5. The dotted line guides the eye.

= (166°, 86°), with a standard deviation of 0.5 Hz to the experimental data (Figure 3A). This compares well with the lowest energy RDC-refined experimental structure's duplex orientation (Θ, Φ) = (164°, 73°).

When the individual slant angles of bond vectors deviate from the ideal values for the A-form or B-form duplex, so does the "RDC wave". Consequently, extracting accurate duplex orientation information (Θ, Φ) from a single type of bond vector becomes difficult. However, nucleic acids possess a number of experimentally accessible dipolar coupling types per turn, and hence the duplex orientation can be determined by averaging the orientations extracted from the "RDC waves" of different bond vector types, such as imino, $C_1'H_1'$, $C_2'H_2'$, $C_3'H_3'$, etc. of the sugars, and C−C and C−H bond vectors of the bases. For example, an average value of (Θ, Φ) = (19°, 58°) was determined using the "RDC waves" from four bond vector types of a simulated RNA duplex with (Θ, Φ) = (20°, 60°). This result was obtained with each bond vector type's slant angles varied randomly by $>\pm 10°$ about their ideal values.

One of the caveats of using RDCs in structure refinement is the degeneracy problem. The solutions to a given RDC are two continuous sets of (θ, ϕ) which lie on two distorted cones about the magnetic Z-axis.[10] However, since the secondary structure information is explicitly embedded in eq 1 and a series of RDCs are simultaneously fit, this degeneracy is reduced to only four distinct possibilities: (Θ, Φ), (Θ, $\Phi+\pi$), ($\pi-\Theta$, $\pi-\Phi$), ($\pi-\Theta$, $2\pi-\Phi$), which yield an identical "RDC wave". The orientation of individual bond vectors in relationship to the alignment tensor axis is uniquely defined for chosen (Θ, Φ). This becomes true because the secondary structure dictates the orientation of individual bond vectors to a uniquely defined orientation and excludes other possible orientations that would otherwise satisfy the RDC values when they are considered individually. This "RDC wave" fitting strategy to remove the degeneracy of individual bond vectors takes advantage of abundant information about the secondary structure which is already available prior to 3D structure calculation. Examples are the chemical shift index, J_{HNHA} coupling, NOE pattern, and NH-exchange data for peptides and the secondary structure with base-pairing schemes and NOE information for nucleic acids. The underlying principle implies that once the duplex orientation is known, the local bond vector orientation can be determined. Although for a given (Θ, Φ) local slant angle variations can result in a distorted "RDC wave, the individual slant angles can be accurately fit (see Supporting Information). For example, the $C_1'H_1'$ RDC bicelle data ($D_a = -16.0$, $R = 0.09$) from the Dickerson DNA dodecamer was fit to determine the viability of determining local conformation from experimental data.[9] First an approximate fit of the helical orientation, (Θ, Φ) = (26°, 125°) was obtained using a fit uniform slant angle of 94°. The individual slant angles were then fit to the experimental data for fixed (Θ, Φ). Since the slant angle fit using the initial duplex orientation parameters yielded a fit with standard deviation of 0.7 Hz, the (Θ, Φ) were individually adjusted by trial and error until the slant angle fit yielded a standard deviation of 0.2 Hz (corresponding to the experimental error[9]). The resultant fit (Θ, Φ) = (23°, 155°) for the RDCs is shown in Figure 3B. The correlation between the fitted slant angles extracted from the "RDC wave" and the slant angles that are calculated from the structure is good, with a Pearson's correlation coefficient, $R = 0.90$. Deviations of the fitted slant angles are attributed to the fact that the Dickerson dodecamer possesses a bend of the duplex axis of 12°/turn.

The RDCs are very sensitive to the orientation of the nucleic acid duplex. Since there are many more residues per turn in nucleic acids than in proteins, detailed features of the "RDC wave" emerge, which contain information of the orientation of both the secondary structure element and individual bond vectors. Furthermore, since eq 1 contains embedded secondary structure information, fits to sets of RDC data result in greatly reduced RDC degeneracy.

Acknowledgment. We thank Dr. Veglia for providing the galley proof of their article prior to publication, and Dr. R. Andrew Byrd for encouragement. The Matlab "RDC wave" fitting codes are available upon request.

Supporting Information Available: Derivation of eq 1, slant angle fit (PDF). This material is available free of charge at http://pubs.acs.org.

References

(1) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. Q. *Rev. Biophys.* **2000**, *33*, 371−424.
(2) Bax, A.; Kontaxis, G.; Tjandra, N. In *Nuclear Magnetic Resonance of Biological Macromolecules*, Pt B; James, T. L., Dötsch, V., Schmitz, U., Eds.; Academic Press: San Diego, 2001; Vol. 339, pp 127−174.
(3) Bondensgaard, K.; Mollova, E. T.; Pardi, A. *Biochemistry* **2002**, *41*, 11532−11542.
(4) Mesleh, M. F.; Lee, S.; Veglia, G.; Thiriot, D. S.; Marassi, F. M.; Opella, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 8928−8935.
(5) Mesleh, M. F.; Opella, S. J. *J. Magn. Reson.* **2003**, *163*, 288−299.
(6) Marassi, F. M. *Biophys. J.* **2001**, *80*, 994−1003.
(7) Mascioni, A.; Veglia, G. *J. Am. Chem. Soc.* **2003**, *125*, 12520−12526.
(8) Mesleh, M. F.; Veglia, G.; DeSilva, T. M.; Marassi, F. M.; Opella, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4206−4207.
(9) Wu, Z. G.; Delaglio, F.; Tjandra, N.; Zhurkin, V. B.; Bax, A. *J. Biomol. NMR* **2003**, *26*, 297−315.
(10) Bax, A. *Protein Sci.* **2003**, *12*, 1−16.

JA039446G