Refinement of Local and Long-Range Structural Order in Theophylline-Binding RNA Using $^{13}C-^{1}H$ Residual Dipolar Couplings and Restrained Molecular Dynamics

Nathalie Sibille,[†] Arthur Pardi,[‡] Jean-Pierre Simorre,[†] and Martin Blackledge*,[†]

Contribution from the Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado, 80309-0125, and Institut de Biologie Structurale, Jean-Pierre Ebel C.N.R.S.-C.E.A., 41 Rue Jules Horowitz, 38027 Grenoble Cedex, France

Received July 6, 2001

Abstract: $^{13}C-^{1}H$ residual dipolar couplings (RDC) have been measured for the bases and sugars in the theophylline-binding RNA aptamer, dissolved in filamentous phage medium, and used to investigate the long-range structural and dynamic behavior of the molecule in the solution state. The orientation dependent RDC provide additional restraints to further refine the overall structure of the RNA−theophylline complex, whose long-range order was poorly defined in the NOE-based structural ensemble. Structure refinement using RDC normally assumes that molecular alignment can be characterized by a single tensor and that the molecule is essentially rigid. To address the validity of this assumption for the complex of interest, we have analyzed distinct domains of the RNA molecule separately, so that local structure and alignment tensors experienced by each region are independently determined. Alignment tensors for the stem regions of the molecule were allowed to float freely during a restrained molecular dynamics structure refinement protocol and found to converge to similar magnitudes. During the second stage of the calculation, a single alignment tensor was thus applied for the whole molecule and an average molecular conformation satisfying all experimental data was determined. Semirigid-body molecular dynamics calculations were used to reorient the refined helical regions to a relative orientation consistent with this alignment tensor, allowing determination of the global conformation of the molecule. Simultaneously, the local structure of the theophylline-binding core of the molecule was refined under the influence of this common tensor. The final ensemble has an average pairwise root mean square deviation of 1.50 ± 0.19 Å taken over all heavy atoms, compared to 3.5 ± 1.1 Å for the ensemble determined without residual dipolar coupling. This study illustrates the importance of considering both the local and long-range nature of RDC when applying these restraints to structure refinements of nucleic acids.

Introduction

NMR spectroscopy is the principal method for determining three-dimensional structure of biomolecules in solution.[1] This technique relies on distance restraints derived from the measurement of cross-relaxation effects (NOE) between protons close in space (<6 Å) supplemented by local dihedral angle information from ^{3}J scalar couplings.[2,3] One limitation of this method is that insufficient distance data, for example at interfacial or hinge regions, can result in ill-defined relative positioning of distant regions or modular domains of the macromolecule. This inability to measure correlated structural information from parts of the molecule which are distant in three-dimensional coordinate space often limits our ability to define the global structure of the molecule. In addition low structural precision due to insufficient constraints cannot easily be distinguished from real dynamic averaging.[4]

Nucleic acids frequently form extended structures; thus, their NMR-derived solution ensembles often exhibit poorly defined long-range order. This is a more serious problem than for proteins, where the density of protons is higher and the more globular structure often supplies sufficient inter-proton NOEs to define the protein fold.[5] The accurate description of the structure and dynamics of nucleic acids in solution is however of fundamental importance for understanding their biological activity. Therefore development of experimental methods that improve long-range structure determination of extended molecules in solution will be particularly useful for the study of nucleic acids.[6]

Recently developed techniques for partially aligning macromolecules in dilute liquid crystal media such as phospholipid bicelles,[7,8] filamentous phage[9,10] or purple membrane frag-

* Corresponding author. Telephone: (33) 4 38 78 95 54. Fax: (33) 4 38 78 54 94. E-mail: martin@rmn.ibs.fr.

† Jean-Pierre Ebel CNRS-CEA.

‡ University of Colorado at Boulder.

(1) Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley: New York, 1986.

(2) Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1997**, *4*, 849−853.

(3) Case, D.; Wright, P. E. In *NMR of proteins*; Clore, G. M., Gronenborn, A. M., Eds.; Macmillan: London, 1993.

(4) Brüschweiler, R.; Case, D. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 27−58.

(5) Varani, G.; Aboud-ela, F.; Allain, F. H.-T. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51−127.

(6) Mollova, E.; Pardi, A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 298−302.

(7) Sanders, C. R.; Hare, B. J.; Howard, K. P.; Prestegard, J. H. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 421−444.

(8) Tjandra, N.; Bax, A. *J. Biomol. NMR* **1997**, *10*, 289−292.

(9) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571−10572.

(10) Hansen, M. R.; Müller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074.

10.1021/ja011646+ CCC: $20.00
Published on Web 11/14/2001

ments[11] have transformed our ability to study long-range order in biomolecules using solution-state NMR. If a macromolecule experiences restricted orientational sampling, for example due to the presence of a dilute liquid crystal, strong first-order interactions are no longer averaged to zero as they are in isotropic solution,[12,13] but instead have a residual component that reports on the average orientation of the interaction relative to the magnetic field. A powerful application of this nonisotropic averaging is the measurement of internuclear residual dipolar couplings (RDCs).[14,15] The inherent strength of the full dipolar coupling interaction allows measurement of RDC under conditions of weak (10^{-3}) alignment, while retaining the rotational diffusion properties necessary for high-resolution solution-state NMR. Assuming the global shape of the molecule is constant, the measured dipolar coupling can be expressed in terms of the orientation $\{\theta, \phi\}$ of the internuclear vector relative to a common alignment tensor for the molecule;

$$D_{ij} = -S\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij}^3}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\varphi\right) \quad (1)$$

where A_a and A_r are the axial and rhombic components of the alignment tensor, r_{ij} is the internuclear distance, and S is the order parameter.[12−15] For dipolar couplings measured between covalently bonded spins where the internuclear distance is fixed, the geometric dependence is purely orientational and provides correlated directional information relative to the molecular coordinate frame.

The orientational information available from RDC is complementary to the standard distance and torsion angle information available from ^{1}H−^{1}H NOE and ^3J coupling, respectively. Nevertheless, structure calculation methods developed for exploiting short-range and local geometric restraints may not be the most appropriate for the determination of molecular conformation using orientational constraints. Thus considerable effort has been invested in the development of novel methods designed to incorporate RDC into structure determination protocols.[16−18] RDC-driven rigid-body molecular modeling methods have been used to define long-range order in extended or multidomain proteins[19−21] and in protein−protein or protein−ligand complexes[22,23] and to determine the relative orientation of helical regions in t-RNA[24] in solution. It has also been shown that experimental RDC can be compared with expected values from structures present in conformational databases, to rapidly identify complete or partial homologous fold[25−29] without recourse to NOE-based structure determination. A recent study has also shown that the complete backbone conformation of a protein can be precisely defined from only RDC, assuming sufficient couplings can be measured from neighboring peptide planes in the presence of two different alignment media.[30]

The introduction of an RDC-dependent term into a hybrid molecular dynamics energy function has also been used to drive an existing structure into a conformation in agreement with measured couplings, NOE data, and known covalent and nonbonding interactions.[31−37] Care must be taken when using this approach to optimally weight the different terms in the potential energy function, for example to avoid unphysical distortions of local or covalent geometry to accommodate the orientational restraints. Other factors which need to be addressed for the optimization of RDC-based molecular dynamics algorithms are the significant orientational degeneracy for a given measured coupling,[38] and the problem that the axial and rhombic components of the alignment tensor (A_a and A_r) and axis orientations (defined by the Euler angles α, β, γ) must be determined prior to, or during, the calculation.[31,33,39]

Residual dipolar couplings also provide otherwise inaccessible information concerning intramolecular and interdomain mobility, due to the dynamic averaging of the dipolar coupling interaction. RDCs are sensitive to motions over a much longer time-scale than those sampled by heteronuclear relaxation measurements, and can therefore extend our understanding of both local and long-range macromolecular dynamics in solution.[19,40−42] The presence of significant internal mobility will also influence RDC-based structure refinement calculations, where it can affect the local order parameter S (eq 1), as well as the assumption that the molecule exists as one single conformation under the influence of a common alignment tensor.

The effects of introducing RDCs into molecular dynamics-based nucleic acid structure refinement protocols have recently been simulated extensively for DNA, showing that RDCs can help define local helical parameters[43] and can determine both local and global features such as DNA bending.[44] Experimental

(11) Sass, J.; Cordier, F.; Hoffman, A.; Cousin, A.; Omichinski, J. G.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047−2055.
(12) Saupe, A.; Englert, G. *Phys. Rev. Lett.* **1963**, *11*, 462−464.
(13) Gayathri, C.; Bothner-by, A. A.; van Zijl, P. C. M.; Maclean, C. *Chem. Phys. Lett.* **1982**, *87*, 192−196.
(14) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(15) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−9283.
(16) Prestegard, J. H. *Nat. Struct. Biol.* **1998**, *5*, 517−522.
(17) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891−5898.
(18) Tjandra, N. *Structure* **1999**, *7*, R205−R211.
(19) Fischer, M. W. F.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013−9022.
(20) Skrynnikov, N.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295*, 1265−1273.
(21) Bewley, C.; Clore, G. M. *J. Am. Chem. Soc.* **2000**, *122*, 6009−6016.
(22) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021−9025.
(23) Bolon, P. J.; Al-Hashimi, H. M.; Prestegard, J. H. *J. Mol. Biol.* **1999**, *293*, 107−115.
(24) Mollova, E.; Hansen, M. R.; Pardi, A. *J. Am. Chem. Soc.* **2000**, *22*, 11561−11562.
(25) Annila, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223−230.
(26) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283−294.
(27) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.
(28) Andrec, M.; Du, P.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222−1229.
(29) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(30) Hus, J.-C.; Marion, D.; Blackledge, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1541−1542.
(31) Tjandra, N.; Omichinski, J.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−738.
(32) Tsui, V.; Zhu, L.; Huang, T.-H.; Wright, P. E.; Case, D. A. *J. Biomol. NMR* **2000**, *16*, 9−21.
(33) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159−162.
(34) Garrett, D. S.; Seol, Y.-J.; Peterkovsky, A.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1999**, *6*, 166−173.
(35) Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M.; Kuszewski. *J. Am. Chem. Soc.* **1999**, *121*, 6513−6514.
(36) Huang, K.; Louis, J. M.; Donaldson, L.; Lim, F.-L.; Sharrocks, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 2615−2628.
(37) Hus, J.-C.; Marion, D.; Blackledge, M. *J. Mol. Biol.* **2000**, *298*, 927−936.
(38) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.
(39) Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77−82.
(40) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nat. Struct. Biol.* **1997**, *4*, 292−297.
(41) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416−1424.
(42) Meiler, W.; Prompers, J.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(43) Trantírek, L.; Urbáek, M.; Tefl, R.; Feigon, J.; Sklenár, V. *J. Am. Chem. Soc.* **2000**, *122*, 10454−10455.

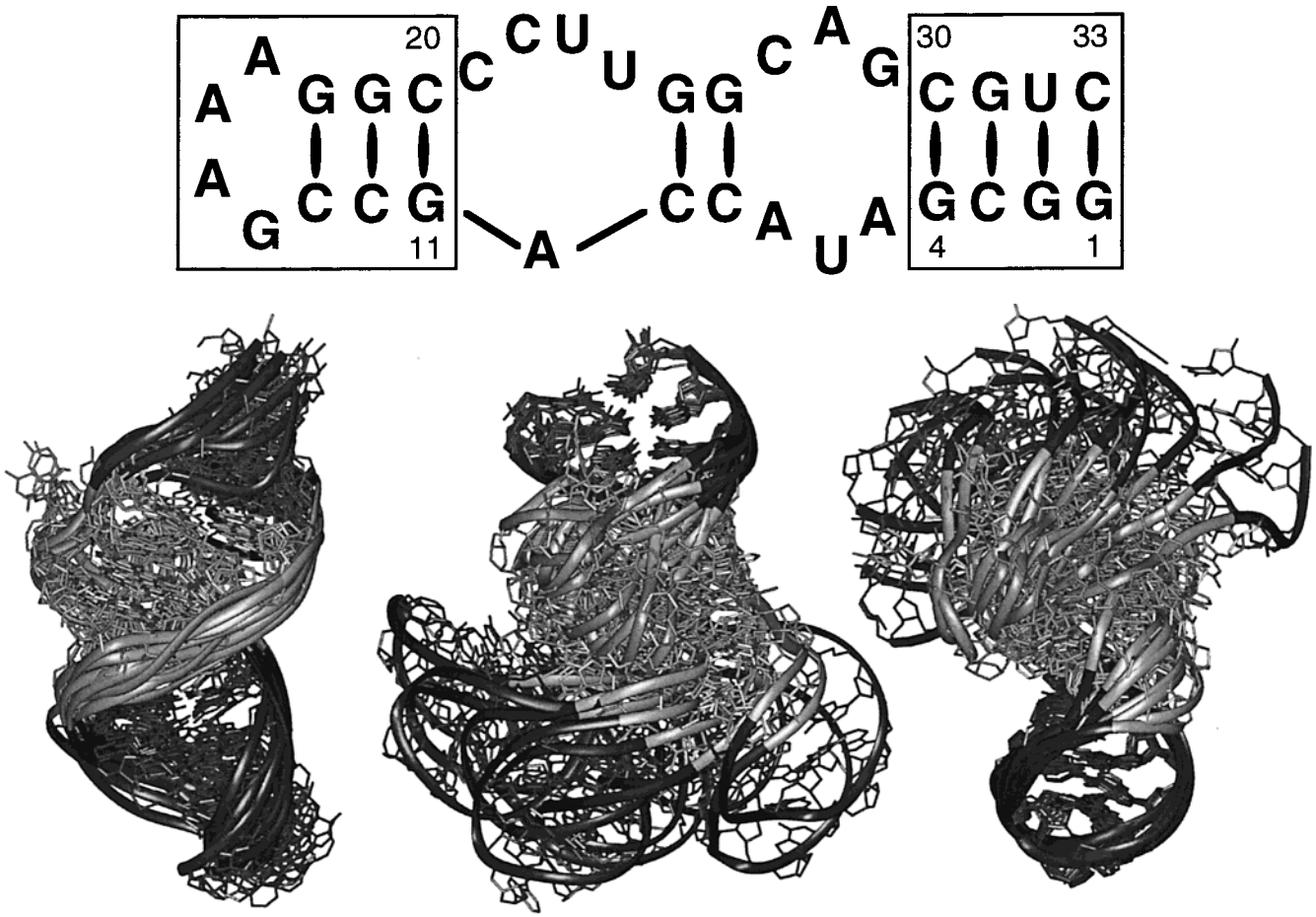

Figure 1. Ensemble of NMR structures determined using NOE and *J*-coupling restraints.[47] (a) Secondary structure of the theophylline binding domain. The stem region I (1−4, 30−33) and the stem II and the GAAA tetraloop region (G11−C20) are shown boxed. (b) Despite well-defined local structure, the long-range order is ill-defined. The shaded regions represent stem I (3−5 terminal stem G1−G4, C30−C33) in dark gray, the theophylline-binding core region (A5−A10, C21−G29, and the theophylline) in light gray and stem II and the GAAA tetraloop (G11−C20) in dark gray (this region is referred to as IIa in the text). (Left) Structures superposed over the entire sequence (1−33), (middle) structures superposed on the average conformation for stem I, and (right) structures superposed on the average conformation for stem II−loop region. All superpositions used only heavy atoms.

studies have also recently demonstrated improved structural precision by combining $^{1}H-^{13}C$, $^{1}H-^{15}N$, or $^{1}H-^{1}H$ RDC measurements with NOE restraints.[45,46]

Here we used RDCs to probe long-range structural order in the theophylline binding aptamer, a 33 nucleic acid RNA consisting of a theophylline binding core region, flanked by a 4 base pair helical domain and a second helical domain terminated with a GAAA tetraloop. The structure of this RNA−theophylline complex in solution was previously calculated by classical NMR techniques, using NOE and *J*-coupling dihedral restraints.[47] The structure was determined precisely enough to identify the mechanism by which RNA recognizes theophylline while discriminating against the structurally similar molecule caffeine. However the NMR ensemble exhibited a much higher level of local than global precision, such that the relative orientation of the two stem regions of the molecule was not well-defined over the ensemble (Figure 1). The study of long-range order of the RNA−theophylline complex in solution using RDCs will thus provide more information on the average global conformation of the molecule. Residual $^{13}C-^{1}H$ dipolar couplings were measured for the RNA−theophylline complex dissolved in aligned Pf1 filamentous phage medium, and these data were used to further refine the local and global structure of the complex. We have attempted to optimize the implementation of RDCs for this application using the molecular dynamics module SCULPTOR,[37] a version of the program Discover. SCULPTOR was written to allow maximal flexibility in the development of conformational search algorithms using long-range structural restraints such as RDCs. In view of the poorly defined long-range order present in the NOE-based ensemble, we have assessed the validity of assuming that molecular alignment can be characterized by a single tensor, by analyzing effective alignment in different regions of the molecule independently. Our approach thus takes into account the local and long-range geometric dependence of RDCs. The results demonstrate substantial improvement in the structural definition of the theophylline complex by inclusion of RDC restraints.

Methods

Sample Preparation. A dilute liquid crystalline sample containing uniformly $^{13}C/^{15}N$-enriched 33-mer RNA and unlabeled theophylline (Sigma) in a molecular ratio of 1:1.1 was prepared in 20 mM phosphate buffer, pH 6.8, 30 mM NaCl, 2 mM $MgCl_2$, and 100% D_2O.[47] The

(44) Vermeulen, A.; Zhou, H.; Pardi, A. *J. Am. Chem. Soc.* **2000**, *122*, 9638−9647.
(45) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 6190−6200.
(46) MacDonald, D.; Herbert, K.; Zhang, X. L.; Polgruto, T.; Lu, P. *J. Mol. Biol.* **2001**, *306*, 1081−1098.
(47) Zimmerman, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J.-P.; Pardi, A. *Nat. Struct. Biol.* **1997**, *4*, 644−649.

liquid crystal consisted of 18 mg/mL of filamentous bacteriophage Pf1.[10] Final sample conditions were 0.75 mM RNA−theophylline in 300 μL, using a thin-wall Shigemi microcell (Shigemi Inc., Alison Park, PA). An identical isotropic sample was prepared as above except without addition of Pf1.[47]

NMR Spectroscopy: Determination of Couplings. All NMR spectra were recorded on a Varian spectrometer operating at a ^{1}H resonance frequency of 600 MHz equipped with a triple resonance, pulsed-field-gradient probehead. All spectra were recorded at 25 °C. Couplings were measured using a 2D ^{13}C−^{1}H TROSY experiment,[48] with spin state selective coherence transfer by gradient[49] to transfer the slowly relaxing component.

Dipolar couplings were derived from the difference in peak splitting in the spectra from the aligned and the isotropic samples. In both cases H$^{(\alpha)}$-TROSY and H$^{(\beta)}$-TROSY experiments were acquired. This allows the measurement of the splitting from the relative positions of the corresponding cross-peaks along the ^{1}H dimension. To help improve the spectral resolution in the ^{13}C dimension, the one-bond carbon−carbon couplings were removed using a constant time (CT) frequency editing period. During this period, TROSY-type spin evolution significantly reduces relaxation-induced signal loss. For the base carbons, the CT-period was set to 16.7 ms and the experiments were recorded as data matrices of 82(^{13}C) × 1024(^{1}H) complex points, with spectral widths of 40.0 ppm (^{13}C) and 13.3 ppm (^{1}H) with a recycling delay of 1.9 s and a total experimental time of 24 h. For the sugar carbons, spectra with a CT-period of 25.0 ms were recorded as data matrices of 132(^{13}C) × 1024(^{1}H) complex points, with spectral widths of 40.0 ppm (^{13}C) and 13.3 ppm (^{1}H) with a recycling delay of 1.9 s and an experimental time of 39 h.

These data were apodized with a 70° shifted squared sine-bell function in the directly detected dimension and a 90° shifted squared sine-bell in the indirectly detected dimension. Data were zero-filled prior to Fourier transformation to final matrices of 2048(^{13}C) × 16384(^{1}H) to help improve digital resolution. The relative peak positions were determined using a Lorentzian line fitting routine available in Felix 2000 (MSI, San Diego, Ca).

In both the aligned and the isotropic state two series of 2D TROSY constant-time [^{13}C,^{1}H] experiments for the bases and sugars were recorded back-to-back. The H$^{(\beta)}$- and H$^{(\alpha)}$-TROSY spectra were recorded in an interleaved manner. Splitting measured from repeated spectra were used to check the reproducibility of the experiments and to ensure that the degree of Pf1 alignment had remained constant during the whole series of measurements in the liquid crystalline state.

Determination of Alignment Tensor Parameters. Tensor eigenvalues and eigenvectors are extracted using a least-squares fitting routine designed to simultaneously determine the five parameters defining the alignment tensor, (A_a, A_r, α, β, γ) by minimizing the residual between all calculated and experimental couplings:[50]

$$\chi^2 = \sum_n \{D_{ij}^{\exp} - D_{ij}^{\text{calc}}\}^2/\sigma_{ij}^2 \qquad (2)$$

σ_{ij} is the estimated uncertainty in the coupling. This least-squares minimization approach is equivalent to the commonly used single value decomposition method.[51] The eigenvalues were ordered such that $|A_{xx}| < |A_{yy}| < |A_{zz}|$, and the Euler angles defining the rotation from molecular to alignment tensor frame were defined such that ($-\pi < \alpha, \gamma < +\pi$) and ($0 < \beta < \pi$).

Structure Calculation Using Orientational Restraints: SCULPTOR. All structure calculations were performed using the program SCULPTOR[37] (Structure Calculation Using Long-range, Paramagnetic, Tensorial and Orientational Restraints)—a suite of in-house programs interfaced to the molecular dynamics program Discover,[52] specifically developed to treat orientation dependent interactions as long-range order restraints.

SCULPTOR allows a flexible approach for exploiting residual dipolar couplings within molecular dynamics-based structure calculations. Tensor parameters are treated as independent pseudo-molecules and are read with the coordinates of the molecule of interest. Eigenvalues and eigenvectors are treated separately, such that one two-point pseudo-molecule represents A_a, a second two-point pseudo-molecule A_r and a third, three-point pseudo-molecule represents the orientation ($A_{\alpha,\beta,\gamma}$) of the tensor. Each term can thus be manipulated independently. The masses of the pseudo-atoms defining the tensorial parameters are empirically set such that the kinetic energy of the molecular system in the temperature range commonly used in restrained MD simulations 300−1000 K is approximately matched to the residual dipolar coupling energy. In this way, a molecular dynamics calculation does not rotate the axis system too rapidly to efficiently anneal to the optimal tensor orientation.

Two independent alignment tensor functions are available in SCULPTOR; these can be used separately for the analysis of restraints measured in different alignment media or, as in this case, to independently study two regions of a single molecule.

Structure Calculation Protocol. Force Field. The program Fdiscover Version 2.98 was modified to include the residual dipolar coupling constraint as an explicit target potential in addition to the classical potential energy function of the AMBER4 force field.[53]

$$E_{\text{RDC}} = k_{\text{RDC}}(D_{ij}^{\text{calc}} - D_{ij}^{\exp})^2/\sigma_{ij}^2 \qquad (3)$$

The uncertainty σ_{ij} was estimated at approximately 0.6 Hz for the measured couplings. To reduce local violations of base geometry that are introduced by the orientational restraints, higher than normal force constants were used for the sites where RDCs were measured. In particular force constants defining base-planarity and local valence angle geometry in the bases were increased.

The quartic repulsion term in the force field used 0.825 of the standard van der Waals radii. The various energetic components were scaled to control the contributions from covalent, nonbonded, and experimental terms during the calculation. No electrostatic terms were used in any of the calculations. All molecular dynamics protocols used weak coupling to a temperature bath[54] unless otherwise stated. Nonbonded interactions used a sigmoid cutoff function, which started at 6.5 Å, such that no interactions beyond 8.0 Å were taken into consideration. Two step sizes were used in the MD calculations—0.1 fs for purely refinement calculations and 0.25 fs for exploratory protocols. These step sizes were used to minimize energetic instability effects, due to the reinforced force field and the steep energy gradients inherent in the RDC target function (eq 2).[55]

(A) Protocol A. The aim of protocol A was not to refine the overall fold of the molecule, but to refine local structure and alignment tensors separately for the two regions stem I (residues 1−4 and 30−33) and the hairpin consisting of stem II and GAAA tetraloop (henceforth referred to as region IIa) as defined in Figure 1. Two independent tensors were used—$\mathbf{A}_1(A_{1a},A_{1r},A_{1\alpha,\beta,\gamma})$ and $\mathbf{A}_2(A_{2a},A_{2r},A_{2\alpha,\beta,\gamma})$ for the two regions. NOE and dihedral angle restraints were used in classical flat-bottomed well potentials for the whole molecule, and no RDCs were used in the calculation for the core region of the molecule. The system was equilibrated at 300 K using direct temperature scaling for 0.5 ps of molecular dynamics and gradually increased to 1000 K over a period of 2 ps, during which time k_{NOE}, k_{dihe}, and k_{RDC} were raised from their initial (0.1, 0.1, 0.001) to their final values (25.0 kcal·mol^{-1}·Å^{-2}, 50.00 kcal·mol^{-1}·deg^{-2}, 1.0 kcal·mol^{-1}·Hz^{-2}). The molecule and tensors then evolved freely for 2 ps before the system was cooled to 100 K over 7 ps and the system minimized using a conjugate gradient algorithm. This calculation was performed using a Lennard-Jones nonbonding potential. To retain the dispersion present in the initial ensemble, 10

(48) Brutscher, B.; Boisbouvier, J.; Pardi, A.; Marion, D.; Simorre, J. P.*J. Am. Chem. Soc.* **1998**, *120*, 11845−11851.
(49) Weigelt, J. *J. Am. Chem. Soc.* **1998**, *120*, 10778−10779.
(50) Dosset, P.; Hus, J.-C.; Marion, D.; Blackledge, M. *J. Biomol. NMR* **2001**, *20*, 223−231.
(51) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334−342.
(52) Molecular Simulations Inc., San Diego.
(53) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.0*; University of California: San Francisco, 1991.
(54) Berendsen, H.; Postma, J. P.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684.
(55) Meiler, J.; Blomberg, N.; Nilges, M.; Griesinger, C. *J. Biomol. NMR* **2000**, *16*, 245−252.

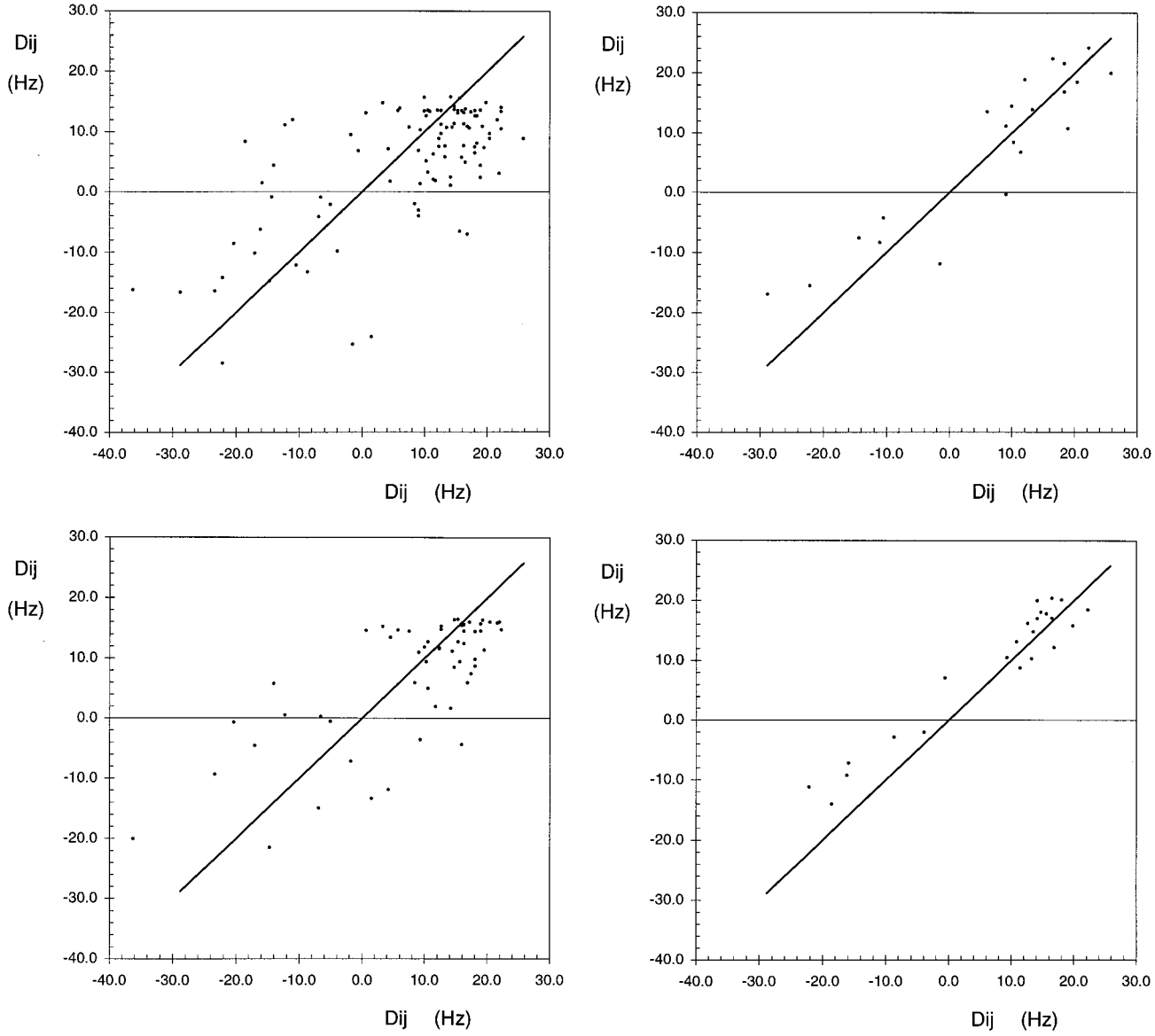

Figure 2. Determination of optimal alignment tensors for the structures in the previously determined NOE-based ensemble.[47] (a) Comparison of all calculated (vertical) and experimental (horizontal) RDCs to the vector orientations in the best-fitting structure from the ensemble of structures. The straight line indicates $D_{ij}^{exp} = D_{ij}^{calc}$. (b) Comparison of calculated (vertical) and experimental (horizontal) RDCs for the fit to canonical (A-form RNA) structure for stem I $\{A_a = 7.3 \times 10^{-4}, A_r = 2.5 \times 10^{-4}\}$, $r = 0.97$. The straight line indicates $D_{ij}^{exp} = D_{ij}^{calc}$. (c) Comparison of calculated (vertical) and experimental (horizontal) RDCs for the best-fitting core structure from the NMR ensemble $\{A_a = 6.8 \times 10^{-4}, A_r = 0.1 \times 10^{-4}\}$, $r = 0.72$. The straight line indicates $D_{ij}^{exp} = D_{ij}^{calc}$. (d) Comparison of calculated (vertical) and experimental (horizontal) RDCs for the fit to canonical (A-form RNA) structure for stem II and the best-fitting loop structure from the NMR ensemble $\{A_a = 7.5 \times 10^{-4}, A_r = 1.1 \times 10^{-4}\}$, $r = 0.93$. The straight line indicates $D_{ij}^{exp} = D_{ij}^{calc}$.

calculations using different initial velocities were performed for each of the 10 initial structures and the lowest energy structure from each of these 10 sets of calculations was used to give the final ensemble of 10 structures.

This set of calculations was repeated using no RDC data, to test for force-field, or protocol dependent effects, and serves as a control for the calculations that included the RDCs.

(B) Protocol B. This calculation used the same NOE and dihedral restraints used in protocol A but also included all the RDC restraints. For each member of the ensemble the local structure of regions stem I and region IIa were tethered to the defined conformation of this structure determined in protocol A in order to facilitate semirigid reorientation of these domains. Distances of less than 6.0 Å were calculated between all carbon atoms in these regions from each structure in the ensemble refined from A. Five percent of the distance was added to and subtracted from the actual value, and these limits were then used as artificial restraints in a standard flat-bottomed potential. The calculation was initiated at 300 K, using direct velocity scaling to equilibrate the system for 1 ps, using a time step of 0.25 fs. The temperature was then increased during the following 3 ps to 1000 K. The molecule and tensor parameters evolved freely for 2.5 ps at this temperature, followed by slow cooling during 7 ps to 50 K and final conjugate gradient minimization. Force constants of $k_{RDC} = 1.0$ kcal·mol^{-1}·Hz^{-2}, $k_{NOE} = 50$ kcal·mol^{-1}·Å^{-2}, and $k_{dihe} = 100$ kcal·mol^{-1}·deg^{-2} were used throughout this calculation. During heating and exploratory periods a quartic nonbonded term was used, which was transformed to the Lennard-Jones potential prior to cooling and minimization.

A set of 10 calculations was performed for each of the 10 structures from the ensemble determined in protocol A. The lowest energy structure from each of these 10 sets of calculations was used to give the final ensemble of 10 structures.

(C) Protocol C. In this protocol each of the structures from protocol B was subjected to a 5 ps low-temperature (300 K) restrained molecular dynamics calculation in the absence of the artificial distance restraints,

using only RDC, NOE, and dihedral angle restraints. k_{RDC} was reduced to 0.025 kcal·mol^{-1}·Hz^2, k_{NOE} to 12.5 kcal·mol^{-1}·Å^{-2}, and k_{dihe} to 25 kcal·mol^{-1}·deg^{-2} during the course of this calculation. The RDC weighting k_{RDC} was empirically set such that the final reduced χ^2/N_{RDC} was approximately equal to 1. This ensemble was taken to represent the final structure in agreement with all experimentally measured data.

(D) Protocol D. This protocol was essentially the same as protocol B except that artificial restraints were not used.

Results and Discussion

A set of 106 ^{13}C−^{1}H residual dipolar couplings was determined for the theophylline binding aptamer, 55 were from the sugars and 51 from the bases (Supporting Information Table S1). To establish how well the measured RDCs fit with the structures generated from classical NMR data, the alignment tensor parameters were determined for each member of the NMR ensemble of structures (Figure 1). All structures in the ensemble have high target functions, and the relative distribution of CH vector orientations of the best-fitting conformation reproduces the experimental data very poorly (Figure 2a). This result is similar to previous observations in experimental and simulation studies of DNA or RNA[10,45] and reemphasizes that the global structure of RNA or DNA is not well-defined by classical NMR data.

We therefore decided to analyze the well-defined domains in the theophylline−RNA complex. On the basis of the observation (presented graphically in Figure 1) that local structural regions are relatively well-defined compared to the overall long-range order,[47] we have attempted to fit local regions to the 5 parameters defining an unknown alignment tensor. This local fitting approach was applied to the stem I, the theophylline binding region, and region IIa, defined as shown in Figure 1 and Methods. The three regions contain 21, 56, and 29 vectors respectively, and thus should allow the 5 parameters defining the alignment tensor to be determined, assuming the coordinates accurately describe the local structure. Dissection of the molecule in this way is clearly somewhat subjective, but this approach was adopted to determine whether data from regions with well-defined local structure were consistent with a single alignment tensor.

The results of this analysis demonstrate that the stem regions provide a more satisfactory fit to the experimental data than the whole molecule or even the theophylline-binding region (Figure 2c). For the regions expected to have canonical structure (stems I and II), we also fit the data to an A-form RNA structure (no NMR restraints were significantly violated—no NOE > 0.2 Å and no dihedral violations > 5°—by assuming A-form structure). The G−U base pair was constructed by replacing the C base with the U base using the graphics program InsightII.[52] In both cases the canonical structures show an improvement in the fit to the RDC data sets compared to the best fit from the NMR ensemble (data not shown). Similar improvement in the quality of the fit has been reported for proteins dissolved in aligned media, when comparing orientational vector distributions with coordinates from crystallographic and NMR models.[40,56,57] This improvement is possibly not surprising, as the orientation of dipolar coupling vectors is essentially defined by measured NOE distances, whose distribution in space is defined by the number of close protons that have well-resolved resonances. This problem may be amplified in the case of nucleic acids, where local motifs, such as a base plane, can be preferentially tilted in one direction because different numbers of NOEs are observed with stacking partners above and below the base of interest. This may be a small effect, and still leave the hydrogen-bonding geometry intact, but the highly sensitive orientational dependence of the measured RDCs could produce a strong residual error term for a vector that deviates slightly from the correct orientation.

Table 1. Structural and Tensorial Statistics from Protocol A: Simultaneous Determination of Local Structure and Alignment Tensor

N^a ensemble	100 RDC	10 RDC	10^b control
E_{phys}[c] (kcal·mol^{-1})	280 ± 29	262 ± 17	235 ± 20
$E_{NOE/dihe}$ (kcal·mol^{-1})	2.86 ± 1.05	2.72 ± 1.12	2.61 ± 0.96
E_{dip} (kcal·mol^{-1})	0.82 ± 0.54	0.76 ± 0.30	
all[d]			
rmsd (Å)		3.5 ± 1.1	3.8 ± 1.1
rmsd[e] (Å)		0.85 ± 0.22	1.25 ± 0.30
stem I		8.86	
A_a (10^{-4})	8.89 ± 0.54	± 0.36	
A_r (10^{-4})	1.17 ± 0.78	1.32 ± 0.72	
rmsd[f] (Å)		1.03 ± 0.26	1.10 ± 0.29
stem IIa			
A_a (10^{-4})	9.56 ± 0.74	9.18 ± 0.13	
A_r (10^{-4})	1.94 ± 0.79	1.88 ± 0.67	

[a] Number of structures taken into account. Statistics are given for the complete ensemble (100) and for the final ensemble of 10 comprising the lowest energy structure refined from each of the initial structures. [b] Calculation performed in the absence of RDC restraints. The original dispersion of stem I, region IIa, and the complete molecule in terms of average pairwise root mean square deviation (rmsd) were 1.51 ± 0.53, 1.11 ± 0.22, and 3.5 ± 1.3Å, respectively. [c] Physical energy of the structure as measured using the reinforced force field. No electrostatic terms were used, and a nonbond interaction cutoff of 8 Å was applied. [d] Average pairwise rmsd was calculated over all heavy atoms. [e] Average pairwise rmsd was calculated over all heavy atoms for stem I (nucleotides 1−4, 30−33). [f] Average pairwise rmsd was calculated using all heavy atoms for stem IIa (nucleotides 11−20).

Average Structures in a Potentially Dynamic System. Interpretation of RDC data in terms of eq 1 assumes that distinct regions of the molecule can be described by the same alignment tensor and therefore experience the same alignment forces. This essentially supposes that the molecule can be considered to be more or less rigid. If the long-range structural disorder present in the NMR ensemble (Figure 1) actually represents the real conformational averaging present in the solution state, then the interpretation of the RDC data as representative of one single conformation under the influence of a common alignment tensor will no longer be valid. Indeed numerous studies have localized different alignment tensors in individual regions of multidomain systems; an observation attributed to the presence of differential dynamic processes in the separate parts of the molecules.[19,58,59] As shown above, fits to the two distinct stem regions of the RNA molecule give similar magnitudes for the alignment tensor eigenvalues. This indicates that similar alignment forces are acting throughout the RNA, although the uncertainties in these values are relatively high given the differences between calculated and experimental couplings (Figure 2b,d). This poor correlation between calculated and experimental values may derive from local structural imprecision, due to the low density of interproton distances in RNA. We have therefore developed a specific restrained molecular dynamics protocol both to address the question of local structural imprecision and to better

(56) Sass, H.-J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303−309.
(57) Tjandra, N.; Marquardt, J.; Clore, G. M. *J. Magn. Reson.* **2000**, *142*, 393−396.
(58) Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1998**, *37*, 116−128.
(59) Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 485−492.

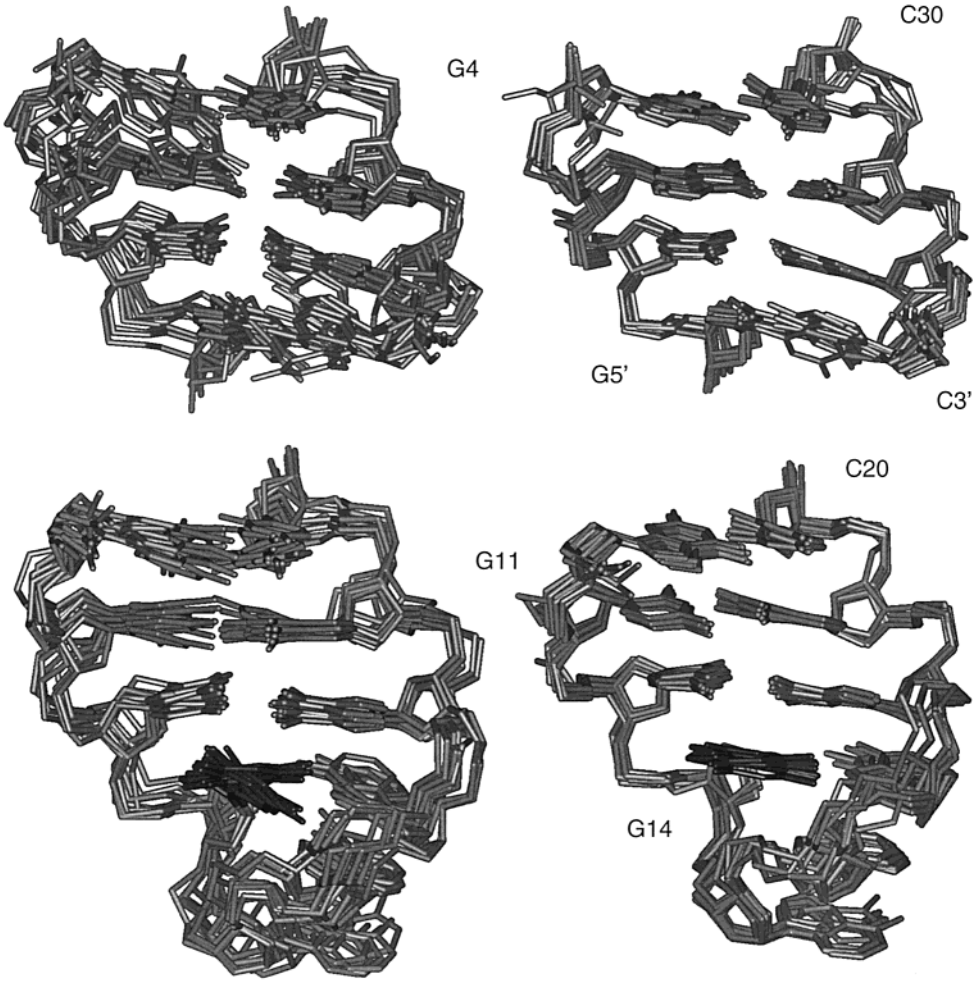

Figure 3. Simultaneous refinement of local structure and determination of alignment tensors using RDCs (protocol A). Restrained molecular dynamics calculation using two independent floating tensors to refine local structure of two distinct regions simultaneously. (a) Local structure in stem I determined using protocol A in the presence (right) and absence (left) of RDCs. (b) Local structure in the region IIa determined using protocol A in the presence (right) and absence (left) of RDCs.

determine the alignment tensor characteristics of the two helical regions of the molecule. RDC restraints were thus applied to the two stem regions (I and IIa) independently such that local structure and alignment tensor characteristics were simultaneously refined.

Determination of Local Structure and Local Alignment Tensors. Recent experiments and simulations have shown that residual dipolar couplings can aid in defining local structure in B-form DNA.[44,45] In these studies, both global and local structure was refined simultaneously using a grid-search approach to optimize alignment tensor values. These results showed that for a relatively well-defined structure such as helical DNA, the minimum in the subspace defined by the interaction parameters (A_a,A_r,α,β,γ) could be determined while refining the helical parameters. Our approach is very similar; the local structures in the two regions stem I and stem IIa, which were already defined from measured NOE and *J*-coupling data, are further refined by the application of RDC restraints. Simultaneously, the alignment tensor eigenvalues and orientation for the two regions were independently determined during the calculation. For reasons of efficiency, we have replaced the grid-search approach by a floating alignment tensor, whose 5 parameters are free to evolve during the structure calculation. This calculation was performed for each member of the ensemble determined using NOE and *J*-coupling restraints.[47] The NOE and *J*-coupling restraints were used for the whole molecule to ensure compatibility between the stem and core structures. To achieve this kind of calculation in SCULPTOR, two independent tensors were defined $\mathbf{A}_1(A_{1a},A_{1r},A_{1\alpha,\beta,\gamma})$ and $\mathbf{A}_2(A_{2a},A_{2r},A_{2\alpha,\beta,\gamma})$; one for each of the identified regions.

The results of this calculation show (Table 1 and Figure 3) that the addition of RDCs refines local structure, particularly in the case of stem I. The disorder present in two regions I and IIa drops from 1.25 ± 0.30 and 1.10 ± 0.29 Å (average pairwise root mean square deviation of the heavy atoms) in the control calculations (no RDC restraints) to 0.85 ± 0.22 and 1.03 ± 0.26 Å in the RDC-restrained version. The stem regions in the control calculations have similar dispersion compared to the previously determined ensemble.[47] Certain bases (for example G14) can be seen to populate multiple conformations that are in agreement with measured NOE in the control ensemble but which are refined to a single orientation by the RDC data (Figure 3). The sugar, backbone and glycosidic dihedral angles are also refined, for example in the central region of stem I (Supporting Information Table S2). The reduced χ^2 for the two regions are 0.11 and 0.09 (yielding effective uncertainty for the experimental RDCs of $\sim$0.20 Hz), which may indicate overconstraint. However, as shown below, the structures are relaxed in the last part of the calculation stage (protocol C) to give a reduced χ^2 of approximately 1 in the final ensemble. The global structural dispersion of the ensemble is essentially unchanged with respect to both control 3.8 ± 1.1 Å ensemble using no RDC constraint

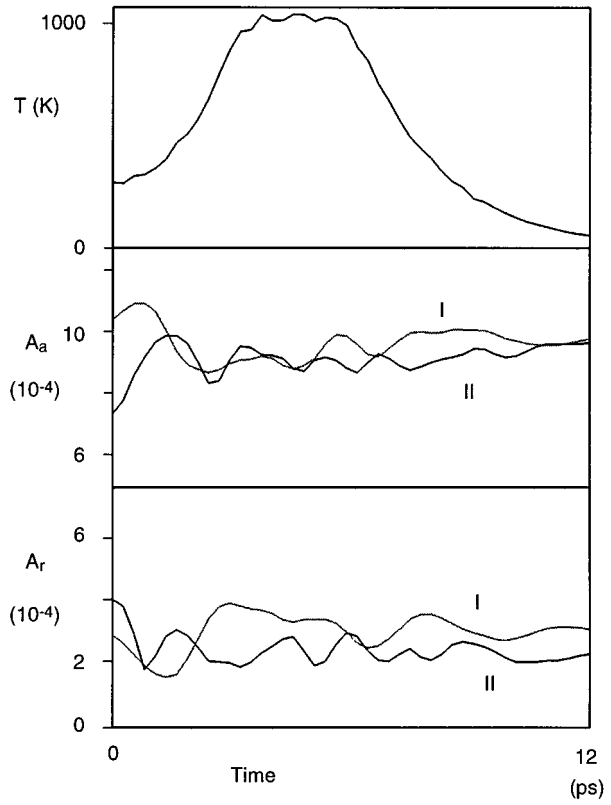

Figure 4. Sampling characteristics of protocol A. Typical evolution of the alignment tensor parameters during the simultaneous tensor/local structure refinement algorithm. (Top) Temperature of the restrained MD calculation, (middle) axial components of the two tensors stem region I (|) and region IIa (||) during the calculation, and (bottom) rhombic components of the two tensors stem region I (|) and region IIa (||) during the calculation.

and the initial 3.5 ± 1.2 Å ensemble. This is not surprising because no long-range order has yet been introduced between the two stem regions.

Importantly the interaction tensors that are determined for the two independently treated regions of the molecule converge to very similar values during the calculation. This is true over the entire ensemble [100 structures calculated − 10 for each initial structure $A_{1a} = (8.89 \pm 0.54) \times 10^{-4}$, $A_{1r} = (1.17 \pm 0.78) \times 10^{-4}$ and $A_{2a} = (9.56 \pm 0.74) \times 10^{-4}$, $A_{2r} = (1.94 \pm 0.79) \times 10^{-4}$ and over the lowest energy ensemble from each of the 10 initial structures $A_{1a} = (8.86 \pm 0.36) \times 10^{-4}$, $A_{1r} = (1.32 \pm 0.72) \times 10^{-4}$ and $A_{2a} = (9.18 \pm 0.13) \times 10^{-4}$, $A_{2r} = (1.88 \pm 0.67) \times 10^{-4}$]. A typical example of the evolution of the alignment tensors during the calculation is shown in Figure 4. This convergence of global interaction parameters and local structural coordinates has important consequences:

First the sets of internuclear vectors for which RDCs are available from each separate region (I and IIa) sample sufficient angular space to determine the alignment characteristics of each region, while simultaneously finding a local conformation in agreement with NOE, *J*-coupling, and RDCs. This was by no means evident a priori and is difficult to predict analytically.[60] The angular dispersion of measured RDCs with respect to the major axis of the determined alignment tensors (A_{zz}) is shown in Figure 5, demonstrating that the internuclear vectors in these two regions (I and IIa) sample a broad region of angular space. It is of interest to note that the largest splitting seen experimentally ($D^{C25}_{C1'-H1'} = -37$ Hz), which is found in the core region, lies within the range of the alignment tensors determined using RDCs from only the stem regions (the largest experimental values for stems I and IIa were −28.8 Hz and +22.2 Hz, respectively).

(60) Fushman, D.; Ghose, R.; Cowburn, D. *J. Am. Chem. Soc.* **2000**, *122*, 10640−10649.

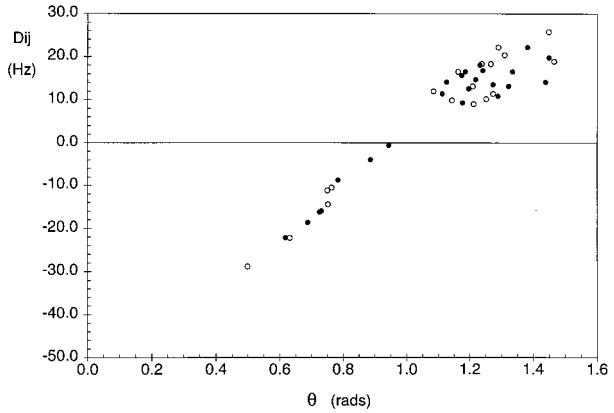

Figure 5. Distribution of the vector orientations with respect to the major axis A_{zz} of the alignment tensors: (open circles) stem I; (filled circles) region IIa, for one of the structures in the locally refined ensemble following calculation A. While the angular dispersion is not complete, most of the range is covered.

Second, this analysis lends stronger support to the previous indication that similar alignment forces are acting on the two distinct regions of the molecule. This implies that the large scale differential domain motion seen in previous studies of extended or multidomain molecules is not taking place here.[19,58,59] While we cannot unambiguously exclude domain motions giving rise to similar apparent alignment characteristics for the different domains,[20] our subsequent analysis will determine the structure of the molecule assuming a common alignment tensor. It should be remembered that this structure will always represent an average conformation in agreement with the all measured NOEs, *J*-couplings, and RDCs.

Semirigid Body Dynamics and Refinement of Core Structure. Each conformer in the ensemble determined in protocol A contains two regions of local structure which have very similar A_a and A_r values for their alignment tensors, although the orientation of the regions is still as disperse as in the original NMR ensemble. A semirigid body dynamics protocol was then performed using a single alignment tensor to drive the global structure into a conformation satisfying all RDC data simultaneously. This was achieved by taking each of the structures in the ensemble A, calculating distances from the locally defined regions I and IIa, and applying these artificial restraints throughout a high-temperature molecular dynamics calculation.[32] In this way the angular coherence, present in each stem region from protocol A, is retained during the calculation and is used to help reorient the entire subdomain of the molecule. Despite the relatively precise definition of the alignment tensors determined during the first step of the calculation, we have chosen to leave the tensor parameters free to evolve toward a mean tensor during protocol B. The RDCs from the core region were also introduced in this step. The aim of this protocol is therefore 2-fold: determination of the overall form of the molecule, driven by the displacement of the two stem regions to a relative orientation in agreement with a common tensor, and refinement of the core region under the influence of this common tensor. The sampling characteristics of the tensorial and Cartesian space are summarized in Figure 6, illustrating the broad reorientational freedom available to the structured domains I and IIa during this calculation.

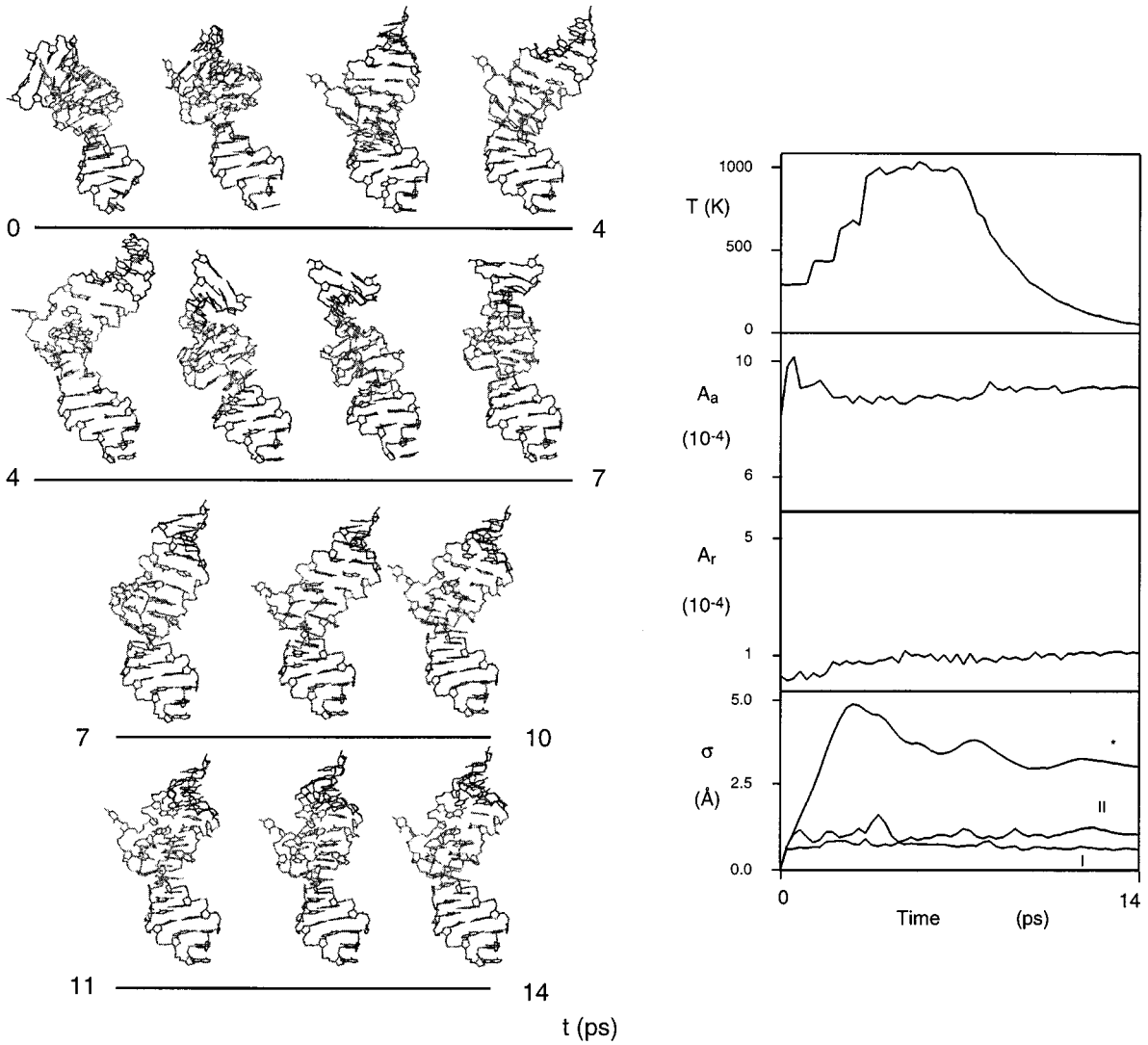

Figure 6. Sampling characteristics of protocol B. Simultaneous refinement of long-range order and local structure. (a) At regular intervals 14 snapshot coordinates were extracted from the restrained MD. Structures were superimposed on the stem IIa region of the molecule (nucleotides 11−20). Note the orientational sampling of the different regions of the molecule. (b) (top) Temperature of the calculation system, (middle) evolution of the eigenvalues of the alignment tensor {A_a and A_a} during the calculation, and (bottom) root mean square deviation of the coordinates of stem I (|) stem IIa (||) and the complete molecule (upper trace) compared to the initial structure.

The results of the calculations for protocol B are summarized in Table 2. The overall order is refined compared to the calculation performed in the absence of the orientational restraints, where the overall root mean square deviation was reduced from 3.8 to 1.5 Å by addition of the RDCs with a common alignment tensor. Not surprisingly the alignment tensor magnitude and directions converge to a well-defined cluster over the whole ensemble. The lowest energy structures from each of the 10 initial conformers were finally refined using a restrained molecular dynamics calculation with the standard AMBER4 force field, using fixed tensors, in the absence of the artificial distance restraints (protocol C). During this calculation k_{RDC} was reduced by a factor of 40, such that the total RDC dispersion present in each structure was approximately equal to the total estimated error. This results in an energetic relaxation of the structure, but no significant change in the dispersion of the structural coordinates. The average reduced χ^2 for the RDC in the final ensemble of structures is 0.95 ± 0.8.

The final ensemble discussed below contains one structure from each of the initial NOE-based conformers (Figure 7, Table 2), and represents the conformation of the molecule in best agreement with all experimentally measured structural restraints. Angular dispersions of measured RDC with respect to the major axis of the alignment tensors (A_{zz}) illustrate the more complete angular sampling due to inclusion of the RDC from the core region (Figure 8).

RDC-Refined Theophylline Binding Aptamer. The stem and tetraloop regions of the molecule retain the structure refined in the initial step of the protocol, while the local structural detail of the core region is refined during steps B and C, due to the introduction of RDC data (Figure 9b). The average pairwise root mean square deviation calculated over all heavy atoms for the core region (nucleotides 5−10, 21−29, and the theophylline) falls from 1.81 ± 0.39 to 1.39 ± 0.22 Å due to the RDC refinement, the most significant effect being observed for the bases. The orientation of G26, A7, and C8, which form one of the base triples, are clearly tilted and refined with respect to the NOE-based structure, while for example A10 and G29 exhibit much lower dispersion due to the RDC refinement. The triple-base hydrogen bonding networks identified in the initial description of this structure[47] are reproduced. The hydrogen bonding geometry stabilizing the theophylline in the core region recognized in the previous study is slightly better defined, due to the higher precision of the surrounding bases—the interactions

Table 2. Structural and Tensorial Statistics from Protocols B−D: Determination of Long-Range Order and Refinement of Local Structure and Alignment Tensor

(a) Energetic and Tensorial Statistics

N^a protocol	10 B	10^b D	10 C
E_{phys} (kcal·mol^{-1})	313 ± 12[c]	507 ± 52	−32.5 ± 5.1[d]
$E_{NOE/dihe}$ (kcal·mol^{-1})	5.0 ± 0.7[e]	11.2 ± 3.2[e]	1.52 ± 0.22[f]
E_{dip} (kcal·mol^{-1})	3.2 ± 0.9[g]	5.5 ± 2.6[g]	2.4 ± 0.2[h]
A_a (10^{-4})	9.28 ± 0.11	9.79 ± 0.90	9.28 ± 0.11
A_r (10^{-4})	1.12 ± 0.14	1.22 ± 0.83	1.12 ± 0.14

(b) Structural Dispersion

rmsd (Å)	no RDC[i]	protocol B	protocol D	protocol C
all[j]	3.8 ± 1.1	1.48 ± 0.22	2.59 ± 0.76	1.50 ± 0.19
core[k]	1.81 ± 0.39	1.37 ± 0.23	1.72 ± 0.35	1.39 ± 0.22
core bases[l]	1.65 ± 0.70	1.29 ± 0.22	1.62 ± 0.57	1.27 ± 0.18
core bb[m]	1.47 ± 0.22	1.33 ± 0.24	1.52 ± 0.15	1.30 ± 0.21
stem I[n]	1.25 ± 0.30	0.96 ± 0.30	1.21 ± 0.24	0.96 ± 0.27
region IIa[o]	1.10 ± 0.29	0.78 ± 0.22	1.61 ± 0.46	0.79 ± 0.18

[a] Number of structures taken into account. Statistic are given for final ensembles of 10 structures comprising the lowest energy structure refined from each of the initial conformers. [b] Calculation performed using direct refinement of the initial ensemble in a one step restrained molecular dynamics in the presence of RDC restraints. [c] Physical energy of the structure as measured using the reinforced force field. No electrostatic terms were used, and a nonbond interaction cutoff of 8 Å was applied. [d] Physical energy of the structure as measured using the standard AMBER4 force field. [e] Calculated assuming $k_{NOE} = 25.0$. [f] Calculated assuming $k_{NOE} = 12.5$ kcal·mol^{-1}·Å^{-2}. [g] Calculated assuming $k_{dip} = 1.0$ kcal·mol^{-1}·Hz^{-2}. [h] Calculated assuming $k_{dip} = 0.025$ kcal·mol^{-1}·Hz^{-2}. [i] This refers to the no-RDC control calculation. Some of these statistics appear in Table 1 and are repeated here for ease of comparison between RDC and non-RDC refined ensembles. [j] Average pairwise rmsd calculated over all heavy atoms. [k] Average pairwise root mean square deviation (rmsd) calculated over all heavy atoms for the core region (nucleotides 5−10, 21−29 and the theophylline). [l] Average pairwise rmsd calculated over all heavy atoms in the base for the core region (nucleotides 5−10, 21−29 and the theophylline). [m] Average pairwise rmsd calculated over all heavy atoms on the phosphodiester backbone core region (nucleotides 5−10, 21−29). [n] Average pairwise rmsd calculated over all heavy atoms for stem I (nucleotides 1−4, 30−33). [o] Average pairwise rmsd calculated using all heavy atoms for region IIa (nucleotides 11−20).

between the theophylline H7, O6, and N9 and the C22 and U24 bases are now populated throughout the ensemble. These are of course relatively small differences, as the local fold of the binding motif was already well-defined in the initial ensemble. It is interesting to note a refinement of the phosphodiester backbone (Figure 9b, Table 2) under the influence of the residual dipolar couplings.

The dispersion of the overall fold of the molecule falls from 3.5 ± 1.2 to 1.5 ± 0.2 Å between the initial and final ensembles, both of which are in agreement with the NOE and *J*-coupling data, underlining the importance of the orientational data available from the RDC measurements. This refinement, and local rearrangement, of the tertiary structure of the core region under the influence of dipolar couplings may explain the poor correlation of the NOE-based core structures with respect to the RDC data noted earlier.

Three-Stage Refinement Using RDCs Compared to Direct Refinement. The three-stage refinement protocol described above was developed for a number of reasons: First we have paid particular attention to the determination of the molecular alignment tensor, due to the importance of this calibration step for the interpretation of RDCs. The alignment characteristics of the molecule were analyzed using individual regions of the molecule based on the observation that local structure was better conserved throughout the NOE ensemble than long-range order. Under certain conditions this approach can also allow the detection of significant differential domain motion in distinct regions of the molecule. In this case the alignment characteristics of the distant stem regions I and IIa were observed to be very similar. Having found no direct evidence for significant motion of these secondary structural motifs, the next step was designed to determine their relative orientation from the available data assuming a single conformation. The local structure in the two domains was also refined during the initial step. This coherent angular information within the two domains was utilized to reorient the modules, which were tethered to their refined local structures during this calculation using artificial restraints. The dispersion in the starting ensemble of structures was retained by applying semirigid reorientation to each member of the ensemble separately. This step also serves to refine the core region under the influence of the tensor, which although still unrestrained, is relatively well-defined by the two coherent domain structures. The final step simply serves to relax the constraints.

By way of comparison we have applied a single algorithm to refine the original NMR ensemble during a high-temperature molecular dynamics calculation under the influence of the RDCs, NOEs, and *J*-couplings. This protocol (D) was identical to protocol B except that no artificial restraints were used to reorient the stem regions and the fact that the initial NOE ensemble was used to provide starting structures. This calculation results in a more diffuse ensemble (average pairwise rmsd of heavy atoms of 2.59 ± 0.76 Å), with higher physical and experimental target functions, although the alignment tensor converges to similar values (Table 2). The reduced χ^2 for the RDCs in this calculation is 0.05 (yielding effective uncertainty for the experimental RDCs of ∼0.13 Hz). The higher covalent energy terms are present due to distortion of the local geometry, particularly of the planarity of the bases. This occurs due to the energetic unfavorability of reorienting entire domains to satisfy the RDCs compared to the lower penalty incurred by local distortion. This illustrates the advantage of refining the local structure and alignment tensor before addressing the global orientation of the different regions of the molecule.

Dynamics of C27 Base Plane. The definition of the alignment tensor allows us to investigate the behavior of the nucleotide C27, which has previously been shown to be dynamic, from the absence of ^{1}H−^{1}H NOE partners for base-plane protons and from the analysis of rotating frame ^{13}C relaxation measurements carried out at the C5, C1′, and C4′ sites.[47,61] The RDC values for the C5−H5 and C6−H6 are also very different from those found in neighboring nucleotides (−12.6 and −4.8 Hz compared to uniformly positive values for all other bases). This supports the observation from the NOE-based structure that this base is not stacked in the core of the molecule. In this case the base plane would be perpendicular to the major axis of the alignment tensor, leading to large positive couplings. In view of the observation that both sugar and base of the C27 are mobile on a μs−ms time scale, these data were excluded from the structure calculations described above. To determine whether any conformations of the nucleotide were in agreement with the measured values for C27, we used the refinement protocol C, with the alignment tensor fixed, and introduced the 5 RDCs measured for C27. As shown in Figure 9, the C27 base can be readily accommodated in a number of different conformations. A detailed analysis of the local dynamic behavior from these data is therefore

(61) Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186−92.

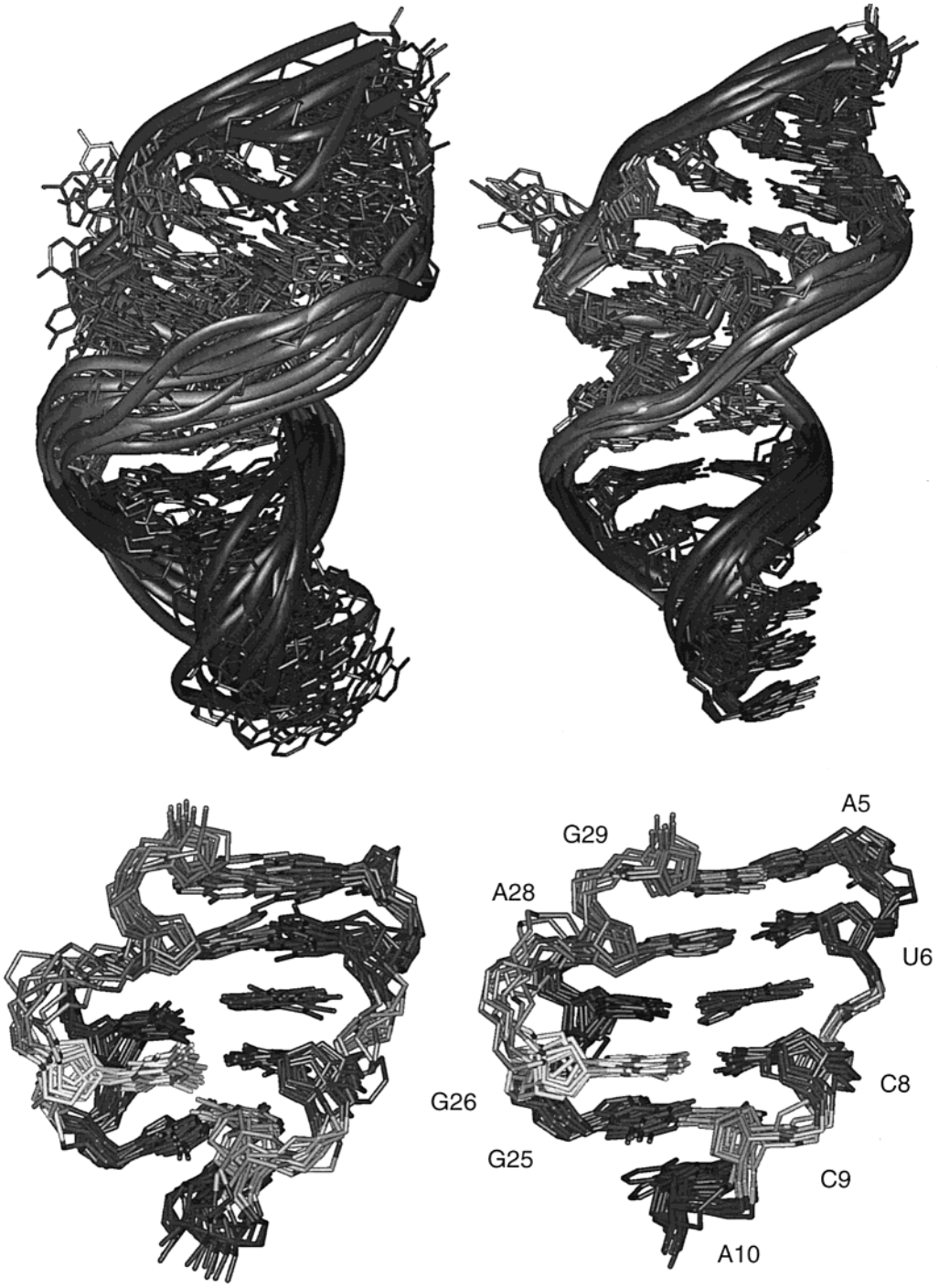

Figure 7. Final ensemble of structures determined using the three-stage protocol. (a) A set of 10 structures comprising the final ensemble. Long-range order is refined (right) compared to the control ensemble (left), which retains the dispersion of the original ensemble. (b) Core region of the final RDC-refined (right) and control (left) ensembles. Note the refinement of orientation of residues A28, G29, A5, and C8 compared to the more disperse definition available in the absence of RDCs. G25, G26, and C9 show a slight reorientation due to the RDC refinement.

beyond the scope of this study, but we note that jumps between any of the conformers shown here cannot be excluded and could give the same RDC values as any one uniquely populated conformer.

These data nevertheless show that isotropic orientational sampling for C27 is unlikely, as this should result in quenched average values for the measured RDC in the base. To illustrate the effects of broad angular sampling on measured RDC, we have calculated the average RDC in the presence of the known tensor during a high-temperature (500 K) restrained molecular dynamics calculation with no restraints on the C27. The coordinates of the rest of the molecule were tethered to their initial conformation, and the alignment tensor was fixed to that determined in the protocol described above (A/B). The RDCs were sampled every 0.1 ps over a 700 ps trajectory. This gives effective RDC values of +4.7 and −1.8 Hz for the C5−H5 and C6−H6 couplings, respectively, illustrating the reduction of effective couplings due to extensive orientational sampling of the base plane. Since the observed RDCs for the base of C27 are larger than the average RDCs in the 700 ps trajectory, this suggests that the base of C27 is more restricted in its motion in solution than the sampling present in this trajectory.

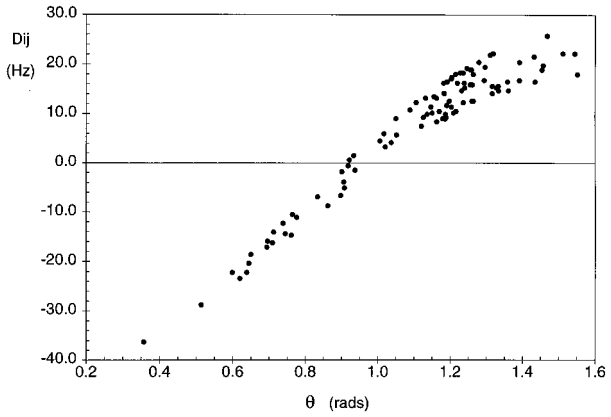

Figure 8. Distribution of the vector orientations with respect to the major axis A_{zz} of the alignment tensor for one of the structures in the final ensemble. By including the theophylline-binding core, the angular sampling is more complete than the distribution of stem I and region IIa (Figure 5).

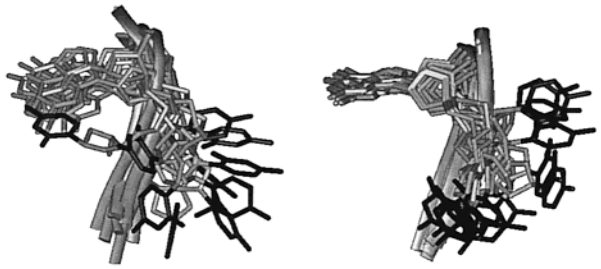

Figure 9. Structural characteristics of the dynamic nucleotide C27 (the base of C27 is shown in black here). Structures calculated using the same protocol in the absence of RDCs (left) and with RDCs for C27 and the rest of the molecule (right). Multiple solutions are available for the orientation of the base in the presence of the RDCs.

Conclusion

We have measured $^{13}C-^{1}H$ residual dipolar couplings from the theophylline−RNA complex, an extended molecular system whose long-range order was poorly defined in the initial NOE-based ensemble. The orientation-dependent RDCs were thus used as additional restraints to further refine the overall structure of the complex. Care must be taken when using RDCs to refine molecular structure in potentially dynamic systems: If the long-range disorder in the initial ensemble represents the real conformational averaging in solution, this mobility may affect the measured RDCs and refinement assuming a single alignment tensor may not be valid. We have therefore developed a specific three-step approach to the application of RDCs to refinement of this structure. Before attempting to determine the average conformation relative to one alignment tensor, we have investigated the alignment characteristics of local substructures of known conformation positioned on either side of the central theophylline-binding region. Restrained molecular dynamics were used to independently determine the optimal alignment tensors of the two helical regions, while simultaneously refining their local structure under the influence of RDC restraints. Very similar alignment tensors were determined for these two distinct regions of the molecule. In the subsequent analysis we therefore assumed that a single molecular alignment tensor is valid for this system. These results do not rule out the presence of large-scale internal motions of different domains of the molecule, which could result in similar effective alignment magnitude for these regions.[20] During this calculation the local dispersion in the helical structures was significantly refined, due to the additional orientational information which complements the previously measured NOE and *J*-coupling data.

In the second step semirigid-body molecular dynamics calculations were used to reorient the refined helical regions to a relative orientation consistent with a common alignment tensor. This step determines the overall form of the molecule, while simultaneously refining local structure in the functionally important theophylline-binding region at the center of the molecule. The overall order and local structure are significantly refined compared to calculations using only NOEs and *J*-couplings. This procedure produces an ensemble of structures whose conformers are in agreement with all experimentally measured restraints and illustrates the importance of considering both local and long-range structure when using RDCs for structure refinement.

Acknowledgment. This work was supported by the Commisariat à l'Energie Atomique and the Centre National de la Recherche Scientifique and with partial support from grants NIH AI 30726and NSF Travel Award INT-96029955 (to A.P.). The authors are indebted to Drs. P. Hanson, S.A. McCallum, H. Zhou, F.M. Jucker, and J.-C.Hus for valuable discussions and K. Wick for sample preparation.

Supporting Information Available: Tables showing residual dipolar couplings and structures refined using protocol A (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA011646+