

RNA unrestrained molecular dynamics ensemble improves agreement with experimental NMR data compared to single static structure: a test case

Robert A. Beckman · David Moreland ·
Shirley Louise-May · Christine Humblet

Received: 23 December 2005 / Accepted: 5 May 2006 / Published online: 28 September 2006
© Springer Science+Business Media B.V. 2006

Abstract Nuclear magnetic resonance (NMR) provides structural and dynamic information reflecting an average, often non-linear, of multiple solution-state conformations. Therefore, a single optimized structure derived from NMR refinement may be misleading if the NMR data actually result from averaging of distinct conformers. It is hypothesized that a conformational ensemble generated by a valid molecular dynamics (MD) simulation should be able to improve agreement with the NMR data set compared with the single optimized starting structure. Using a model system consist-

ing of two sequence-related self-complementary ribonucleotide octamers for which NMR data was available, 0.3 ns particle mesh Ewald MD simulations were performed in the AMBER force field in the presence of explicit water and counterions. Agreement of the averaged properties of the molecular dynamics ensembles with NMR data such as homonuclear proton nuclear Overhauser effect (NOE)-based distance constraints, homonuclear proton and heteronuclear ^1H - ^{31}P coupling constant (J) data, and qualitative NMR information on hydrogen bond occupancy, was systematically assessed. Despite the short length of the simulation, the ensemble generated from it agreed with the NMR experimental constraints more completely than the single optimized NMR structure. This suggests that short unrestrained MD simulations may be of utility in interpreting NMR results. As expected, a 0.5 ns simulation utilizing a distance dependent dielectric did not improve agreement with the NMR data, consistent with its inferior exploration of conformational space as assessed by 2-D RMSD plots. Thus, ability to rapidly improve agreement with NMR constraints may be a sensitive diagnostic of the MD methods themselves.

R. A. Beckman (✉) · D. Moreland · S. Louise-May ·
C. Humblet
Discovery Research Informatics, Computer-Assisted Drug
Design, Pfizer Global Research and Development, Ann
Arbor, MI 48105, USA
e-mail: eniac1@snip.net

R. A. Beckman
Biophysics Research Division, University of Michigan, Ann
Arbor, MI 48109-1055, USA

R. A. Beckman
Division of Pediatric Hematology/Oncology, University of
Michigan Medical Center, Ann Arbor, MI 48109-0238,
USA

Present Addresses:

R. A. Beckman
Clinical Research & Development, Hematology &
Oncology, Centocor, Inc., 200 Great Valley Parkway,
Malvern, PA 19355-1307, USA

S. Louise-May
Neurogen Corporation, Branford, CT 06405, USA

C. Humblet
Wyeth Research, Princeton, NJ 08543, USA

Keywords Molecular dynamics · Nuclear magnetic resonance · Particle mesh Ewald · Ribonucleic acid

Abbreviations

RNA	ribonucleic acid
PME	particle mesh Ewald
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
2D	two-dimensional
RMSD	root mean square deviation
ps	picoseconds

fs	femtoseconds
Å	angstroms
K	Kelvin
°C	degrees Centigrade
kcal	kilocalories
mol	mole
AMBER	assisted model building with energy refinement
ISPA	isolated spin pair approximation
IRMA	iterative relaxation matrix analysis
T_1	longitudinal NMR relaxation time
T_2	transverse NMR relaxation time
J	NMR coupling constant
NOESY-JRE	nuclear Overhauser effect spectroscopy, with jump return excitation

Introduction

Nuclear magnetic resonance (NMR) spectroscopy provides structural and dynamic information on biological macromolecules. However, NMR experimental observables are averages, often non-linear, of multiple conformational states and dynamic modes. Therefore, single optimized solution structures derived from refinement against NMR constraints may be misleading if the underlying data actually results from the averaging of multiple distinct conformers.

Molecular dynamics (MD) simulations have already demonstrated that alternative conformers which are different than the single optimized NMR structures can be broadly consistent with experimental NMR constraints, for both peptides [1] and RNA [2]. We hypothesized, however, that a conformational ensemble generated by a valid MD simulation should be able to *improve* agreement with the NMR data set compared with the single optimized starting structure.

MD, restrained by NMR constraints, has long been utilized to refine single static NMR structures [3]. Distances inferred from the nuclear Overhauser effect (NOEs) and dihedral angles derived from NMR coupling constants (J) are typically part of the restraining function, and agreement is often systematically quantified with a penalty function or R factor. Time averaging has also been taken into account in restrained simulations, where the restraints are time averaged, and the R factor is for an ensemble rather than for an individual structure [4–8].

Dynamics from free MD simulations have also been compared to NMR data. Model-free [9] and spectral

density-based [10] treatments allow for comparison of molecular motion models with heteronuclear NMR data from labeled macromolecules, and these methods have been applied to proteins [11, 12] and nucleic acids [13]. However, heteronuclear NMR data require ^{15}N and/or ^{13}C labeling of macromolecules and as such this data may not be routinely available for all systems of interest.

Structural ensembles from free MD have been compared to experimental structures, both for nucleosides [14] and DNA [15]. These studies must deal with the ambiguity of inferring experimental structures from the underlying experimental constraints [14].

Free MD has also been compared directly to experimental NMR constraints such as NOEs and coupling constants. In an MD study of structural contributions to binding energies in drug–DNA complexes, Gallego et al. [16], used comparison to NMR constraints for the purpose of cross validation of the simulation results. An extensive and systematic comparison of a long (50 ns) peptide MD simulation to experimental constraints was performed by Daura et al. [1]. They concluded that conformers dramatically different from the single NMR structure could be included while only slightly worsening the agreement with the NMR data. However, at no point in the long simulation did agreement with the NMR data improve.

We now report free dynamics ensembles for RNA with improved agreement with experimental NMR constraints, compared with the single highly refined, average NMR structure. The MD simulations were part of an exploratory methods development effort aimed at utilizing free MD for this purpose. They were brief (300 ps) and it is likely that longer simulations and methodologic improvements could have led to further performance enhancement. The generality of this result requires further investigation, and even in the test case here, the improvement was more notable for one RNA octamer, compared to a related one where improvements were limited to improvements in agreement with coupling constant constraints.

A negative control simulation (500 ps) with damped molecular motions did not lead to similar improvements. The performance of the primary simulation and the negative control, when judged by agreement with NOE data and homonuclear and natural abundance heteronuclear (^1H – ^{31}P) coupling constant (J) data, mirrored similar differences in two-dimensional root mean square deviation (2D RMSD) plots, frequently used as a measure of the degree to which MD explores conformational space [17–19]. This suggests the possibility that the ability to improve agreement with NMR data may also be a sensitive diagnostic in evaluating MD simulations.

Methods

Model system

As a model system we used two related RNA sequences shown by homonuclear proton NMR and heteronuclear ^1H - ^{31}P NMR to adopt different conformations at central tandem G-A mismatches [20, 21]. Specifically, the sequence (rGGCGAGCC)₂ adopts a “sheared” conformation with a narrow minor groove and an increased helical twist between the tandem mismatches. In contrast, (rGCGGACGC)₂ adopts an “imino” conformation with imino hydrogen bonds between A N1 and G NH1, a widened minor groove, and decreased helical twist between tandem mismatches (Fig. 1).

Tandem G-A mismatches are extremely common in RNA, occurring in ribosomal RNA, and both the DNA/RNA and RNA hammerhead ribozyme [22–24]. The context, and in particular the 5′ Watson-Crick base pair, can affect structural features of sheared base pairs such as sugar pucker and stabilizing hydrogen bonds

[25], and can also affect the choice between sheared and imino hydrogen bonding schemes [22, 26, 27].

The system we study in this paper was of interest due to the availability of high resolution NMR structures based on data (primarily homonuclear) from unlabeled RNA, and the conformational and hydrogen bonding variation displayed by the two closely related sequences.

Dynamics

Dynamics utilized the AMBER 4.1 force field [28, 29] with explicit water and counterions. The starting structures were derived from the average refined NMR structures, minimized and relaxed in the presence of the water and counterions.

The most significant difference between the primary simulations (300 ps) and the negative control simulations (500 ps) was in the approximation of the electrostatic component of the force field. The primary simulations utilized the particle mesh Ewald (PME) method, which has been applied with success to the

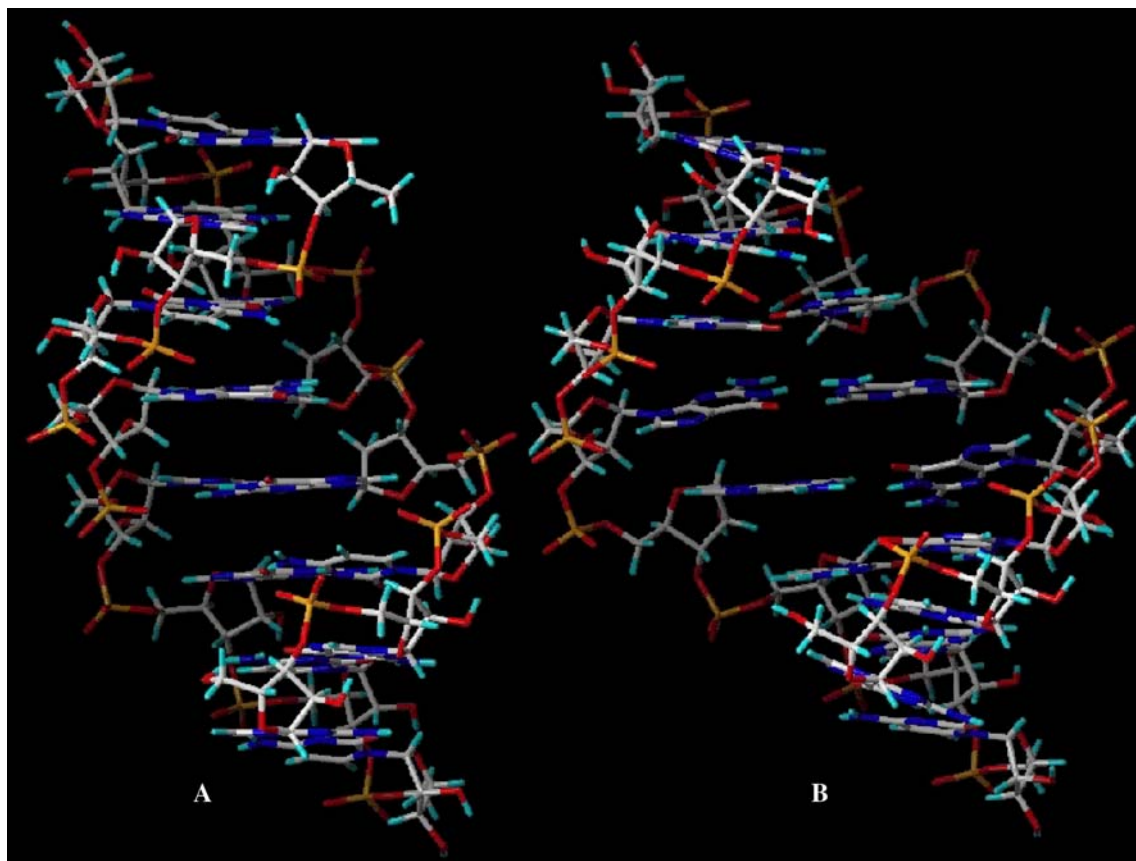


Fig. 1 Solution structures for (rGGCGAGCC)₂ and (rGCGGACGC)₂ as determined by NMR spectroscopy. Carbon atoms are shown in white, hydrogen in light blue, nitrogen in dark blue, oxygen in red, and phosphorus in yellow

simulation of proteins and nucleic acids in water subject to periodic boundary conditions [30–32]. The negative control simulations utilized a distance-dependent dielectric with explicit waters, a technique no longer in general use which is expected to damp molecular motions. Guenot and Kollman [33, 34] have shown in protein simulations with explicit water that structural results with a distance dependent dielectric (DDD) are not dependent on the presence or absence of a non-bonded cut-off, and that such simulations are energetically stable and free of solute–solvent temperature differential and solvent evaporation artifacts, but protein motions were attenuated indicating limited exploration of conformational space. In addition, the distance dependent dielectric method assumes spherical screening which may be inappropriate for helical macromolecules.

The primary and negative control simulation methodologies were optimized separately, not in a parallel fashion. Therefore, the comparison cannot be viewed as a rigorous comparison of PME and DDD electrostatics, but rather a comparison of these specific simulation methods which differed in other respects as well.

The simulations were brief exploratory simulations, and should not be viewed as definitive. Details of the simulation methodologies are given in the Appendix.

NMR R factors

R factors were defined to quantify the agreement of the original NMR structures (R_{NMR}) and the ensembles of structures created by molecular dynamics runs (R_{md}) with homonuclear nuclear Overhauser effect (NOE) derived distance constraints, and homonuclear (HCCH) and heteronuclear (HCOP) coupling constants (J). The NMR structures were single average optimized structures from an ensemble generated by NMR refinement including NMR constrained molecular dynamics in vacuo.

The R factor for agreement of the original NMR structures with NOE based distance constraints, $R_{\text{NMR-NOE}}$, is calculated as follows:

$$R_{\text{NMR-NOE}} = (1/n_{\text{distance constraints}}) \sum_{\text{distance constraints}} (|r_{\text{struct}} - r_{\text{nmr}}|)/(r_{\text{struct}} + r_{\text{nmr}}), \quad (1)$$

where $n_{\text{distance constraints}}$ is the total number of NOE based distance constraints, including “non-NOE” lower bound constraints (i.e. lower distance bounds due to absence of an NOE); r_{struct} is the distance in

the NMR structure between the atoms subject to a particular NOE constraint; r_{nmr} is the midpoint of the NOE-based distance constraint for conventional constraints and the lower bound for “non-NOE constraints”; and, for non-NOE based constraints, a term is added to the numerator above only if $r_{\text{struct}} < r_{\text{nmr}}$.

Similarly, the R factor R_{md} , for the agreement of the ensemble of structures generated by dynamics, with NOE based distance constraints, is given by

$$R_{\text{md-NOE}} = (1/n_{\text{distance constraints}}) \sum_{\text{distance constraints}} (|r_{\text{md}} - r_{\text{nmr}}|)/(r_{\text{md}} + r_{\text{nmr}}), \quad (2)$$

where r_{md} is the $1/r^6$ weighted average of the distances between these two atoms in the ensemble generated by the dynamics run:

$$r_{\text{md}} = \left\{ \sum_{\text{structures}} [(1/r^6)]/n_{\text{structures}} \right\}^{-1/6}. \quad (3)$$

In (3), $n_{\text{structures}}$ is the number of structures included in the ensemble, and in (2), “non-NOE” based lower bounds are treated as they are in (1).

A supplementary analysis was also carried out using analogous expressions, but calculating NOE R -factors using the full constraint breadth utilized for NMR structure refinement, which involved no penalty within the constraint range and a penalty outside this range equal to the absolute value of the difference between the observed distance and the nearest bound (i.e. upper or lower) divided by the sum of the constraint midpoint and the observed distance.

The R factor for the agreement of the NMR derived structures with the coupling constant (J) data, $R_{\text{NMR-J}}$, is given by:

$$R_{\text{NMR-J}} = (1/n_{\text{coupling constant constraints}}) \sum_{\text{coupling constant constraints}} (|J_{\text{struct}} - J_{\text{nmr}}|)/(J_{\text{struct}} + J_{\text{nmr}}), \quad (4)$$

where $n_{\text{coupling constant constraints}}$ is the total number of NMR based coupling constant constraints, the J_{nmr} are the NMR derived coupling constants, and the J_{struct} are those determined from the NMR structure, using the expressions [35]:

$$J_{\text{HCOP}}(\text{Hz}) = 15.2 \cos^2 \theta - 6.1 \cos \theta + 1.6 \quad (5)$$

and

$$J_{\text{HCCH}}(\text{Hz}) = 10.2 \cos^2 \theta - 0.8 \cos \theta \quad (6)$$

where the θ represent the relevant dihedral angles for phosphodiester bonds and for ribose sugar protons, respectively.

Expressions analogous to (4) are used to calculate R factors for the ensemble of structures generated by dynamics, $R_{\text{md}-J}$, with the coupling constant data, wherein J_{struct} is replaced by J_{md} . J_{md} are calculated from (5) or (6) as appropriate, using the linear average of all J s calculated from each structure in the ensemble for J_{md} :

$$J_{\text{md}} = \left(\sum_{\text{structures}} J \right) / n_{\text{structures}}. \quad (7)$$

Separate R factors, $R_{\text{NMR}-J\text{-HCCH}}$ and $R_{\text{MD}-J\text{-HCCH}}$ were calculated for the sugar couplings (H5' and H5'' excluded due to the lack of stereospecific assignments) and corresponding R factors, $R_{\text{NMR}-J\text{-HCOP}}$ and $R_{\text{MD}-J\text{-HCOP}}$, for the phosphodiester couplings.

Overall R factors, $R_{\text{NMR}-J\text{-overall}}$ and $R_{\text{MD}-J\text{-overall}}$, were calculated for all coupling constants from the weighted average of the separate J coupling R factors noted above, weighted by the number of constraints in each R factor.

Overall structure R factors, $R_{\text{NMR-overall}}$ and $R_{\text{MD-overall}}$, were calculated from the weighted average of the R factors for all constraint classes, again weighted by the number of constraints in the class. This analysis was also performed for the supplemental NOE analysis utilizing the full NMR constraint breadth.

Hydrogen bond occupancy

Hydrogen bond occupancies were determined according to Lindauer et al. [36], who use two distances and three angles in their determinations. The atoms in question are H, the hydrogen; D the hydrogen bond donor atom (bonded to H); and A the hydrogen bond acceptor, bonded to C, which is either a carbonyl carbon if A is a carbonyl oxygen, or the midpoint of two carbons C1 and C2 if A is a heteroaromatic nitrogen. The criteria for a hydrogen bond are as follows:

$$\begin{aligned} \text{Distance A-H} &< 2.5 \text{ \AA} \\ \text{Distance A-D} &< 3.9 \text{ \AA} \\ \text{Angle D-H-A} &> 90^\circ \\ \text{Angle H-A-C} &> 90^\circ \\ \text{Angle D-A-C} &> 90^\circ. \end{aligned} \quad (8)$$

2-D RMSD plots

The behavior of the molecular structures during simulation was examined using 2D-RMSD plots. The root-mean square difference in atom positions was determined for all pairwise combinations of conformations that were sampled during the simulation. For the primary (PME) simulations the RMSD values were determined with CARNAL. For the negative control (DDD) simulations, RMSD values were determined with a Sybyl Programming Language script that used the Biopolymer Fit command. The values for each simulation were tabulated in separate files. The files were read into the SAS program and 2-dimensional contour plots were generated with PROC GCONTOUR.

Results

Energy stability

The primary (PME) simulations gave stable dynamics trajectories. The negative control (DDD) simulations, in contrast, displayed a slow decrease in energy, amounting to approximately 8%, including sudden drops at approximately 300 ps (sheared conformer) and 190 ps (imino conformer) before finally equilibrating after 400 ps (sheared conformer) and 300 ps (imino conformer). There were no solvent evaporation artifacts or vacuum bubbles in any of the simulations. Energy plots for the trajectories are shown in Fig. 2A, B (negative control simulation, DDD) and 2C, D (primary simulation, PME). Severely unstable trajectories, RNA unwinding, and solvent evaporation artifacts were observed in related simulations without PME and with a constant dielectric of 1 (data not shown).

Agreement with the NOE-based NMR distance constraints

The R factors describing the agreement of the NMR starting structures and the molecular dynamics ensembles with the NOE derived NMR based distance constraints, for both primary (PME) and negative control (DDD) simulations are given in Table 1A. About 174 constraints were used for $r(\text{GGCGAGCC})_2$, 209 constraints for $r(\text{GCGGACGC})_2$.

We note that the NMR structure for the “sheared” conformer $r(\text{GGCGAGCC})_2$ agrees less completely with the NOE data which determined it than does the

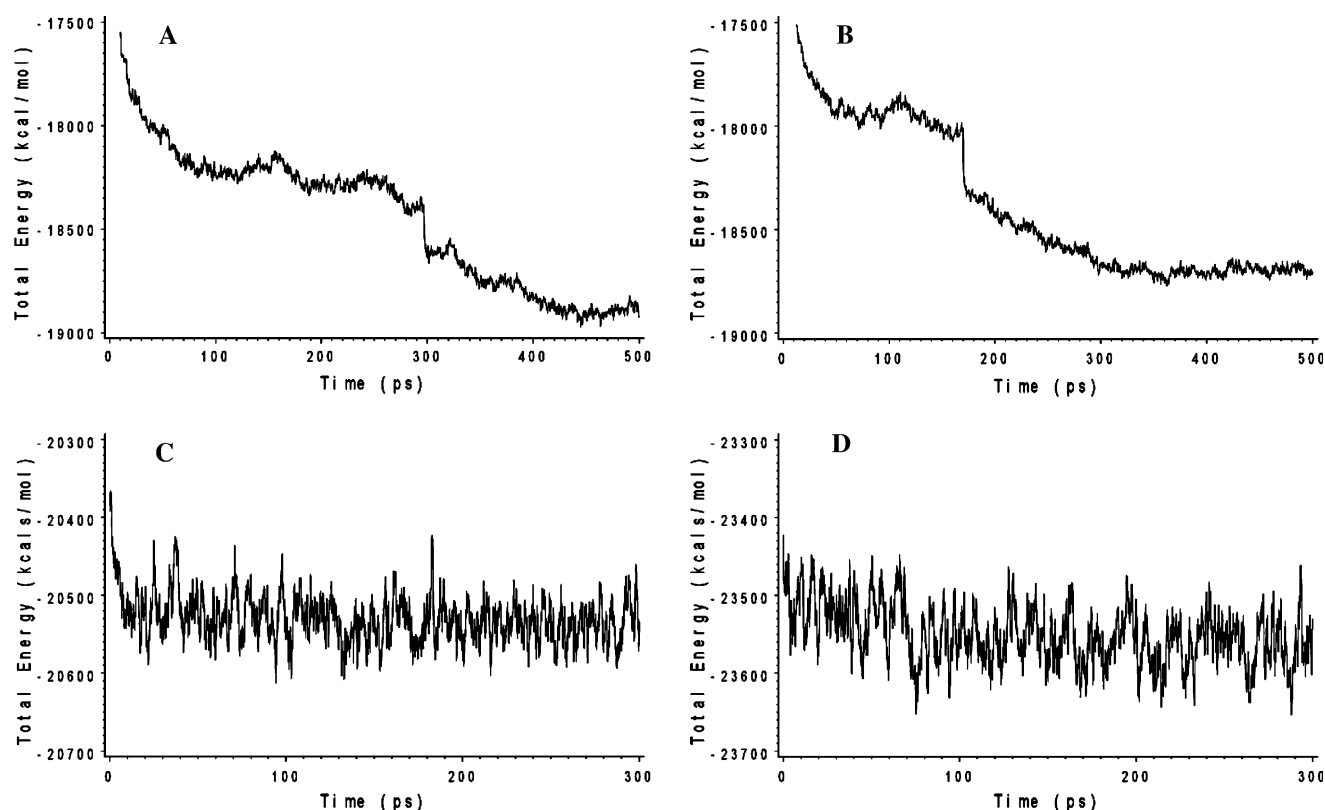


Fig. 2 Energy plots for (rGGCGAGCC)₂ (**A** and **C**) and (rGCGGACGC)₂ (**B** and **D**) simulations for the negative control simulations (**A** and **B**) and primary simulations (**C** and **D**)

“imino” conformer r(GCGGACGC)₂. The primary (PME) simulation leads to a sheared conformer ensemble with a far better *R* factor than the original NMR starting structure, and an imino conformer ensemble which is comparable in agreement to the NMR starting structure. In contrast, the negative control (DDD) simulation shows no improvement for the sheared conformer and a dramatic worsening for the imino conformer compared to the NMR starting structures.

Agreement with NMR coupling constant-based dihedral angle constraints

The *R* factors describing the agreement of the NMR starting structures and the molecular dynamics

ensembles with the coupling constant derived NMR-based dihedral angle constraints, for both primary (PME) and negative control (DDD) simulations are given in Tables 2 (sugar dihedrals) and 3 (phosphodiester dihedrals).

For the sugar dihedrals, 14 constraints were available for the sheared conformer r(GGCGAGCC)₂ and 17 constraints for the imino conformer r(GCGGACGC)₂. Agreement of the NMR structures with these constraints was not as complete as for the NOE-based distance constraints, and the level of agreement was similar for the two conformers. Both ensembles from the primary (PME) simulations had improved agreement compared to the NMR starting structure. Results compared to the NMR structures were similar for the negative control (DDD) simulation.

Table 1 *R* factors determined for H-H distance constraints, (A) main analysis using midpoints of NMR constraints (B) Supplemental analysis using full breadth of NMR constraints

Octamer	Number of constraints	<i>R</i> _{NMR-NOE}	<i>R</i> _{md-NOE} Primary simulation	<i>R</i> _{md-NOE} Negative control simulation
A.				
(rGGCGAGCC) ₂	174	0.088	0.052	0.090
(rGCGGACGC) ₂	209	0.048	0.053	0.129
B.				
(rGGCGAGCC) ₂	174	0.021	0.002	0.025
(rGCGGACGC) ₂	209	0.020	0.027	0.100

Table 2 *R* factors determined for H-H dihedral angle constraints

Octamer	Number of constraints	$R_{\text{NMR-}J\text{-HCCH}}$	$R_{\text{md-}J\text{-HCCH}}$ Primary simulation	$R_{\text{md-}J\text{-HCCH}}$ Negative control simulation
(rGGCGAGCC) ₂	14	0.187	0.142	0.148
(rGCGGACGC) ₂	17	0.186	0.104	0.108

Table 3 *R* factors determined for H-P dihedral angle constraints

Octamer	Number of constraints	$R_{\text{NMR-}J\text{-HCOP}}$	$R_{\text{md-}J\text{-HCOP}}$ Primary simulation	$R_{\text{md-}J\text{-HCOP}}$ Negative control simulation
(rGGCGAGCC) ₂	7	0.128	0.161	0.168
(rGCGGACGC) ₂	7	0.218	0.173	0.205

For the phosphodiester dihedrals, 7 constraints were available for the sheared conformer r(GGCGAGCC)₂ and 7 constraints for the imino conformer r(GCGGACGC)₂. The NMR structures had only fair agreement with these constraints, worse for the imino conformer than the sheared. Compared to the NMR structures, the primary (PME) simulation worsened the *R* factor for the sheared conformer and improved it for the imino. Results for the negative control (DDD) simulation were similar, producing *R* factors which were slightly inferior to those achieved in the primary simulation.

Overall agreement with coupling constant constraints

The overall agreement of the NMR structures and MD ensembles with all coupling constant constraints is given in Table 4. Both primary and negative control simulations improved overall agreement with coupling constant constraints, the primary simulation to a slightly greater extent.

Overall agreement with total NMR constraint set

The overall agreement of the NMR structure and MD ensembles with all NMR NOE and coupling constant constraints is given in Table 5A for the primary analysis using NOE constraint midpoints, and in Table 5B for the secondary analysis using full NOE constraint breadths. The primary simulation resulted in approximately 36% improved agreement for the sheared conformer r(GGCGAGCC)₂ as judged by the primary analysis, or a 57%

improvement as judged by the secondary analysis, while leaving the agreement for the imino conformer r(GCGGACGC)₂ unchanged. The negative control simulation had little effect on the agreement for the sheared conformer r(GGCGAGCC)₂, while worsening the agreement for the for the imino conformer r(GCGGACGC)₂ by 106% as judged by the primary analysis, or by 174% as judged by the secondary analysis.

Hydrogen bond occupancy

Hydrogen bond occupancies for the sheared conformer r(GGCGAGCC)₂ and the imino conformer r(GCGGACGC)₂ are given in Tables 6 and 7, respectively. Hydrogen bond occupancies cannot be assumed to be fully equilibrated in an MD simulation, as the timescale for their opening and closing is frequently on the order of milliseconds, far exceeding the timescale provided by MD (see Discussion).

Of note, in the primary simulation for the sheared conformer the adenine (A13) amino proton to guanine (G4) N3 hydrogen bond, and the symmetry-related bond between the A5 amino proton and G12 N3 are partially and symmetrically occupied, 68.5% and 71.3%, respectively, consistent with equilibration of the simulation across the symmetry axis despite the brief simulation time. In the negative control simulation these bonds are 75.1% and 99.7% occupied respectively, indicative of a lack of equilibration across the same symmetry axis.

In the imino conformer, the G7 imino hydrogen bond is 79.6% occupied in the primary simulation, in approximate agreement with its estimated linewidth

Table 4 Overall weighted average *R* factors over all coupling constant constraints

Octamer	Number of constraints	$R_{\text{NMR-}J\text{-overall}}$	$R_{\text{md-}J\text{-overall}}$ Primary simulation	$R_{\text{md-}J\text{-overall}}$ Negative control simulation
(rGGCGAGCC) ₂	21	0.167	0.148	0.155
(rGCGGACGC) ₂	24	0.195	0.124	0.136

Table 5 *R* factors determined for total NMR constraint set, (A) main analysis using midpoints of NOE constraints, (B) supplemental analysis using full breadth of NOE constraints

Octamer	Number of constraints	<i>R</i> _{NMR-overall}	<i>R</i> _{md-overall} Primary simulation	<i>R</i> _{md-overall} Negative control simulation
A.				
(rGGCGAGCC) ₂	195	0.097	0.062	0.097
(rGCGGACGC) ₂	233	0.063	0.060	0.130
B.				
Octamer	Number of constraints	<i>R</i> _{NMR-NOE}	<i>R</i> _{md-NOE} Primary simulation	<i>R</i> _{md-NOE} Negative control simulation
(rGGCGAGCC) ₂	195	0.037	0.016	0.039
(rGCGGACGC) ₂	233	0.038	0.037	0.104

Table 6 Hydrogen bond occupancies for r (GGCGAGCC)₂

Base Pair	Donor	Acceptor	Occupancy (%) Primary simulation	Occupancy (%) Negative control simulation
1	G1-NH ₂	C16-O2	100.0	99.8
1	G1-NH	C16-N3	100.0	100.0
1	C16-NH ₂	G1-O6	97.0	99.0
2	G2-NH ₂	C15-O2	100.0	99.2
2	G2-NH	C15-N3	100.0	100.0
2	C15-NH ₂	G2-O6	99.5	99.8
3	G14-NH ₂	C3-O2	99.8	100.0
3	G14-NH	C3-N3	100.0	100.0
3	C3-NH ₂	G14-O6	99.4	100.0
4	G4-NH ₂	A13-N7	99.8	100.0
4	A13-NH ₂	G4-N3	68.5	75.1
5	G12-NH ₂	A5-N7	100.0	100.0
5	A5-NH ₂	G12-N3	71.3	99.7
6	G6-NH ₂	C11-O2	99.7	100.0
6	G6-NH	C11-N3	100.0	100.0
6	C11-NH ₂	G6-O6	99.2	100.0
7	G10-NH ₂	C7-O2	100.0	99.8
7	G10-NH	C7-N3	0.0	100.0
7	C7-NH ₂	G10-O6	99.4	100.0
8	G9-NH ₂	C8-O2	100.0	99.2
8	G9-NH	C8-N3	100.0	99.8
8	C8-NH ₂	G9-O6	97.3	99.3

from [21]. This dynamics is not reproduced in the negative control simulation.

However, both simulations show a high level (>98%) of occupancy in the imino conformer of the G4 imino mismatched hydrogen bond, which is not consistent with its linewidth indicating only partial occupancy.

It is difficult to interpret the significance of terminal fraying or the lack of it, given the terminal hydrogen bond constraints which were required in the primary but not the secondary simulation. Experimentally, terminal fraying in both conformers is indicated by solvent exchange broadening with the G1 imino and in the imino conformer by its solvent exchange peaks in the NOESY-JRE spectrum.

This limited fraying is not reproduced in the simulations of the sheared conformer, likely due to the applied hydrogen bond constraints in the primary

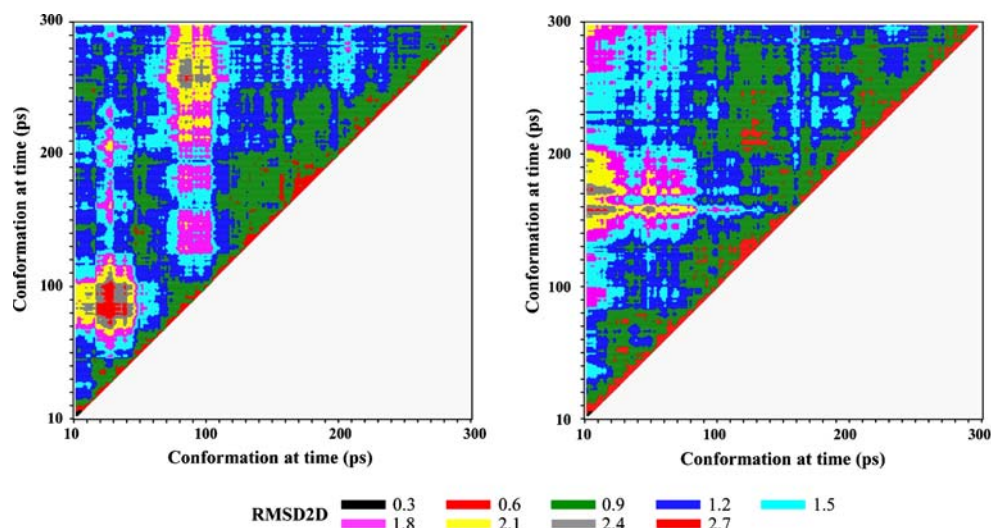
simulation, and to incomplete exploration of conformational space in the negative control simulation. For the imino conformer, terminal hydrogen bonds such as G1 amino to C16 carbonyl and G9 amino to C8 carbonyl are less fully occupied in the primary simulation than the negative control simulation. Thus, dynamics is seen in the primary simulation even in base pair hydrogen bonds which were artificially constrained. Both the artificial constraints and the short simulation time preclude detailed analysis.

2D-RMSD Plots

Two-dimensional RMSD plots are shown for the primary (PME) simulations in Fig. 3A, B, and for the negative control (DDD) simulations in Fig. 4A, B. The primary (PME) simulations show significant variation

Table 7 Hydrogen bond occupancies for r(GCGGACGC)₂

Base Pair	Donor	Acceptor	Occupancy (%) Primary simulation	Occupancy (%) Negative control simulation
1	G1-NH ₂	C16-O2	78.8	86.7
1	G1-NH	C16-N3	100.0	98.4
1	C16-NH ₂	G1-O6	99.2	99.4
2	G15-NH ₂	C2-O2	100.0	99.2
2	G15-NH	C2-N3	100.0	100.0
2	C2-NH ₂	G15-O6	99.5	100.0
3	G3-NH ₂	C14-O2	99.5	99.2
3	G3-NH	C14-N3	100.0	100.0
3	C14-NH ₂	G3-O6	94.9	95.6
4	G4-NH	A13-N1	98.6	98.6
4	A13-NH ₂	G4-O6	98.6	94.0
5	G12-NH	A5-N1	93.0	100.0
5	A5-NH ₂	G12-O6	99.2	99.6
6	G11-NH ₂	C6-O2	97.8	99.8
6	G11-NH	C6-N3	99.8	100.0
6	C6-NH ₂	G11-O6	96.0	99.8
7	G7-NH ₂	C10-O2	99.7	100.0
7	G7-NH	C10-N3	79.6	100.0
7	C10-NH ₂	G7-O6	78.0	100.0
8	G9-NH ₂	C8-O2	57.9	100.0
8	G9-NH	C8-N3	99.8	100.0
8	C8-NH ₂	G9-O6	99.8	99.2

Fig. 3 Two-dimensional RMSD plots for (rGGCGAGCC)₂ (A) and (rGCGGACGC)₂ (B) primary (PME) simulations

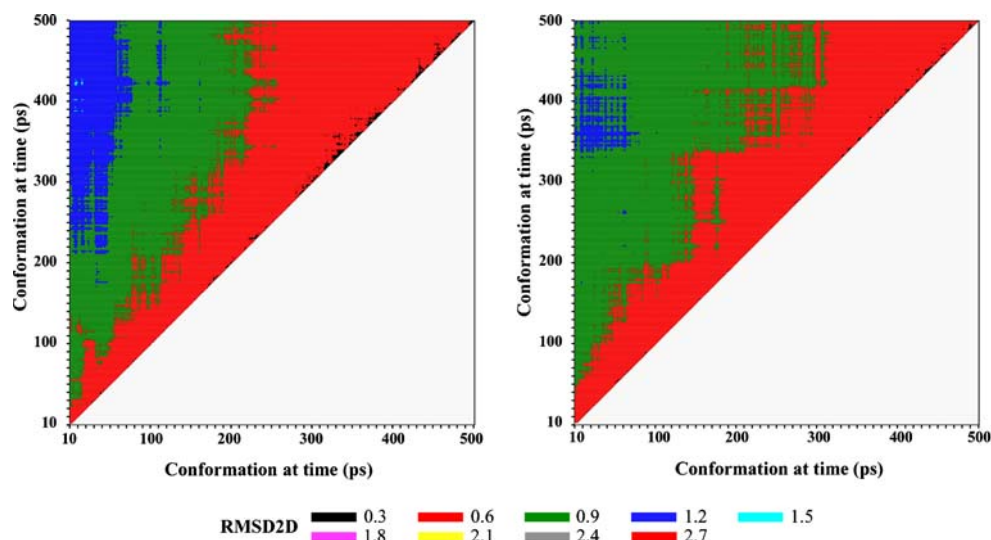
in RMSD, earlier exploration of very different conformations and occasional return to structures more like the original within the time of the simulation, despite its brevity. In contrast, the results from the negative control (DDD) simulation demonstrate a gradual drift from the starting structure. Upon energy equilibration, even this conformational drift slows down.

Discussion

The NMR observables of homonuclear NOE and homonuclear and heteronuclear coupling constants are

the product of conformational averaging in the test tube (and *non-linear* averaging in the case of the NOE), and generation of a valid ensemble of structures from a molecular dynamics trajectory would therefore be expected to enhance agreement with this NMR data. At a minimum, a valid MD trajectory should produce structures that are actually populated in the test tube, and by dramatically increasing the number of degrees of freedom, could enhance agreement with experimental data. In turn, significant worsening of agreement would suggest that the trajectory explored structures that do not actually exist, reflecting either insufficient simulation time for equilibration, or defects

Fig. 4 Two-dimensional RMSD plots for (rGGCGAGCC)₂ (**A**) and (rGCGGACGC)₂ (**B**) negative control (DDD) simulations



in MD methodology. We investigated a test case of the ability of free MD to generate conformational ensembles with enhanced agreement with NMR constraints compared to a single static structure, utilizing a primary (PME) simulation and a negative control (DDD) simulation, in which damping of molecular motion was expected.

R-factors

Time-dependent build-up of homonuclear ^1H – ^1H NOEs can be modeled more precisely by including the effects of local motion and anisotropic tumbling from molecular dynamics simulations [37, 38]. However, we have utilized a simpler approach involving the NOE-based distance constraints themselves, compared to observed $1/r^6$ -weighted distance averages from the molecular dynamics simulations. The rate of build-up of NOEs is generally proportional to a $1/r^6$ -weighted distance average between the two protons. However, for fast motions such as methyl rotation, faster than the overall molecular tumbling, averaging may be $1/r^3$ weighted [39]. The effect of employing this latter weighting was not evaluated in this paper. In a comparison between the two weightings in a peptide system, better agreement was seen with $1/r^6$ weighting than with $1/r^3$ weighting [1].

Our approach utilizing NOE-based distance constraints also has pitfalls. In particular the two structures were refined using distance constraints which were derived from primary NOE data using different methodologies: the “sheared” conformer refinement utilized the isolated spin pair approximation (ISPA), which ignores spin diffusion, whereas the “imino” conformer refinement utilized iterative relaxation

matrix analysis (IRMA) [40], which iteratively accounts for spin diffusion within the network of assigned spins. Since the breadth of distance constraints determined by the two methods will not be equivalent, and since the breadth of NMR distance constraints is often arbitrary, we chose not to use the whole breadth of the distance constraint in the R factors. Our choice of the midpoint of the constraint as the target for the R factor controls for the breadth of the constraints, but is in itself arbitrary and involves small R factor penalties for structures that may actually be consistent with the NMR data (Tables 1A and 5A). We also carried out a supplementary analysis (Tables 1B and 5B), calculating NOE R -factors using the full constraint breadth, which involved no penalty within the constraint range and a penalty outside this range equal to the absolute value of the difference between the observed distance and the nearest bound (i.e. upper or lower) divided by the sum of the observed distance and the constraint midpoint. The trends and conclusions drawn from this supplemental analysis were exactly those of the main analysis. However, compared to the main analysis, the absolute R factors did decrease as predicted by providing a penalty free region within the constraint bounds, particularly with the broad distance constraints resulting from the ISPA treatment.

The experimental coupling constants are point estimates from the NMR spectra, and the expected NMR refined and MD coupling constants can be calculated from the structures, as described above. However, the number of coupling constant constraints is limited and therefore they alone may not provide a statistically valid diagnostic by which to compare MD trajectories with experimental data.

The R factors describing agreement of the NMR structures with the NMR constraints demonstrate several trends. First, the R factors show better agreement with the NOE data than with the coupling constant data, an expected result given the large number of NOE constraints (compared to coupling constant constraints) that were used in the NMR structure refinement. The R factor shows better agreement with the NOE data for the imino conformer, which was refined using IRMA, thereby accounting for spin diffusion, compared with the sheared conformer, in which the ISPA assumption ignores spin diffusion. This suggests utility of the IRMA method. Nonetheless, the $R_{\text{NMR-NOE}}$ is still low for the sheared conformer, suggesting that the effect of spin diffusion is limited. This is consistent with the long T_1 relaxation times observed for this molecule.

With regard to the dynamics trajectories, the R factors for the primary (PME) and negative control (DDD) simulations demonstrate differences, with superior performance of the primary simulation as expected. Overall performance of both simulations is better for the sheared conformer than for the imino conformer, and this may be due to the fact that the latter has been more highly refined through the use of IRMA. Alternatively, the fact that performance of the primary simulation is better for the sheared conformer than for the imino conformer may reflect a lack of complete generality of this methodology as a structure refinement tool. The generality of the methodology is still unknown and can only be determined by application to a large number and variety of test cases. In particular, we would anticipate that the methodology would have limited utility in situations where the key molecular motions which limit agreement with the NMR data are on a timescale which is comparable to or exceeds that of the MD simulation.

The primary simulation achieved a significant improvement in the NOE-based R factor for the sheared conformer and demonstrated no change for the imino conformer (Table 1A and B). The R factor for sugar dihedrals improved for both conformers (Table 2). The R factor for phosphodiester dihedrals worsened considerably for the sheared conformer, while improving for the imino conformer (Table 3). This might represent a statistical artifact given the small numbers of constraints involved (7), although it could also indicate difficulties of the simulation around highly charged centers, or the need for a longer simulation time for equilibration. Overall, agreement with all coupling constant data was improved for both conformers (Table 4), and agreement with the total constraint set was substantially improved for the

sheared conformer while remaining unchanged for the imino conformer (Table 5A and B).

The negative control simulation showed no change in the NOE-based R factor for the sheared conformer, but led to significant worsening for the imino conformer (Table 1A and B). The R factor for sugar dihedrals improved for both conformers (Table 2). The R factor for phosphodiester dihedrals worsened considerably for the sheared conformer, while improving slightly for the imino conformer (Table 3). Overall, agreement with all coupling constant data was improved for both conformers, although only slightly for the sheared conformer (Table 4), and agreement with the total constraint set remained unchanged for the sheared conformer while worsening substantially for the imino conformer (Table 5A and B).

Hydrogen bond occupancy

NMR data sets also give qualitative constraints. In particular, chemical shifts and linewidths of base protons determine the likely degree to which they occupy hydrogen bonds with potential base-pairing partners. Agreement of simulations with these constraints cannot consistently be expected as equilibration of hydrogen bond opening and closing may occur on a millisecond timescale. In some cases, hydrogen bonds have been seen to form in approximately 2 nanoseconds in MD simulations of RNA [2].

The intrinsic chemical shift for a typical Watson-Crick hydrogen-bonded guanine imino proton, obtained from a combination of NMR measurements and ring current effect calculations from nearest neighbor base pairs is 13.5 ppm [41–43], compared to the solvent shift of 4.9 ppm which would be expected for a proton in full exchange with solvent. The observed chemical shifts of guanine imino protons in the conformers may represent a linear average between these two states if the exchange is rapid on the NMR time scale. However, due to specific electrostatic shielding and deshielding and ring current effects which depend on the sequence and on the precise tertiary structure, one must consider these calculations very approximate. Moreover, if the hydrogen bonded protons are in slow or intermediate exchange on the NMR time scale (the more likely case in the exchangeable proton NMR experiments), the percent occupancy cannot be estimated by averaging chemical shifts.

In the case of slow or intermediate exchange one can use the linewidths to estimate the hydrogen bond occupancy, since exchange with solvent will broaden the NMR lines, enabling calculation of the exchange rate. Fully exposed imino protons at pH 6.4 (the

conditions of the exchangeable proton spectra for the sheared conformer) exchange at a rate of approximately 500 per second at 25 °C [44].

For the G1 imino proton of the sheared conformer, a linewidth of approximately 25 Hz is estimated from Fig. 6 of [20]. Given the much narrower intrinsic linewidths on the order of 2–3 Hz, most of the 25 Hz linewidth is due to exchange broadening, indicating an exchange rate of 2π times this or approximately 160 per second, 30% of that of a fully exposed imino proton. Hence, up to 30% of the terminal imino protons may be exposed even at 20 °C. In contrast, the G2 imino linewidth appears similar to its intrinsic linewidth of 2–3 Hz, indicating exchange broadening of 1 Hz or less, an exchange rate of less than 2π per second, and hence probably greater than 98% occupancy. The G3 imino resonance is intermediate between these two cases but estimation of the linewidth is difficult due to overlap with G2. It would appear that the G3 imino hydrogen bond is at least 85% occupied from this analysis. Within the mismatches, while the optimized NMR structures derived from the overall constraint set indicate atypical hydrogen bonds between the guanine and adenine nucleotides, the data do not require that these be fully occupied in a dynamic model. The relevant A5 and G4 amino protons are only tentatively assigned, and thus the hydrogen bond occupancy cannot be estimated from the NMR data.

For the imino conformation, chemical shift and NOE data also indicate that the three guanine imino protons are involved in Watson-Crick hydrogen bonds. The broadening of the G1H1 resonance in the one-dimensional imino proton spectrum as the temperature is raised from 0° indicates some terminal fraying at 40 °C, on the order of 30–60%, as estimated from Fig. 2 of [21]. In a NOESY-JRE spectrum [45, 46] at 15 °C, all four guanine imino protons show some exchange with water, suggesting some degree of fraying, but this is greater for the terminal G1H1 and the mismatched G4H1 (also on the order of 30–60% exposed by estimation of linewidth), less for the G3 and G7 iminos. This is consistent with the narrower linewidths of the G3 and G7 iminos, which as estimated from Fig. 2 of [21], indicate 98–100% and approximately 80% occupancy of their respective hydrogen bonds.

Slightly broadened NMR lines at 10 °C are seen for the mismatched imino hydrogens in a thermodynamic study of related RNA mismatches [26], again consistent with partial occupancy. Subsequent heteronuclear NMR data using a J_{NN} HNN-COSY experiment confirms the mismatch hydrogen bonding in another related sequence, but again full occupancy of these bonds is not required by the data [27]. Figure 2C, D of

this paper show that the crosspeaks corresponding to the mismatch hydrogen bonds are less in volume than those corresponding to Watson-Crick base pairs, which themselves may not be fully occupied, suggesting reduced occupancy, although this volume decrease may also be due to effects of the limited ^{15}N bandwidth. Furthermore, the experiments are at low temperature (10 °C) in 4 mM MgCl_2 , which would tend to increase hydrogen bond occupancy.

In the primary simulations, overall hydrogen bond architecture was appropriately preserved, and occupancy of symmetry-related hydrogen bonds was similar, despite the short length of the simulation. The starting structures had symmetrical and complete occupancy of these hydrogen bonds, but dynamics was seen in the simulation and was comparable on either side of the symmetry axis. In other studies with asymmetrical RNA starting structures, asymmetry has been preserved for the full 16 ns of simulation [2].

The fraying seen in the imino conformer primary simulations is consistent with the overall NMR data set, which indicates some terminal fraying of the hydrogen bonds in question. The hydrogen bond occupancy predicted for the G7 imino proton in the imino conformer by the simulation is consistent with the NMR linewidth of that proton. The occupancy of terminal G1 imino hydrogen bonds is overestimated, likely due to the terminal hydrogen bond constraints applied. For the imino conformer, the G4 imino mismatch hydrogen bond is 98.6% occupied in the simulation, while its NMR linewidth and NOESY-JRE exchange rate suggest much more dynamics, i.e. 30–60% fraying. Thus while the dynamics of the G7 imino is reproduced, the greater dynamics of the G4 imino, as indicated by the NMR data, is not. However, longer simulations are likely required to equilibrate hydrogen bonds internally in polynucleotide duplexes.

In the negative control simulation the asymmetric occupancy of symmetry-related mismatch hydrogen bonds (G4-A13 and A5-G12) in this simulation suggests a failure to maintain equilibration across the symmetry axis despite the symmetrical starting structure.

The negative control simulation demonstrates full occupancy of terminal hydrogen bonds in the sheared conformer, not consistent with the NMR data, which indicates some degree of terminal fraying. This may be due to the overall dampening of molecular motion or failure to equilibrate within the simulation time. For the imino conformer, the order of dynamics of imino protons shown by the NOESY-JRE data is preserved: i.e., the G3 and G7 imino hydrogen bonds are more fully occupied than those involving the G1 and G4 iminos. However, the degree of occupancy of all

hydrogen bonds is over-estimated compared to the NMR data, with the exception of the G3 imino, for which near full occupancy is suggested by the line-widths. Again this suggests damping of molecular motion or failure of equilibration in the simulation time.

2D-RMSD Plots

2D-RMSD plots are also important diagnostics of the ability of a simulation to explore conformational space and support the observations from the R-factor analyses above. One would expect structures to eventually not merely drift from a starting point, but to revisit initial families of conformations, providing a varied profile in the 2D-RMSD plots. The simulation length required for revisiting of initial families of conformations is unknown, but even in the short simulations evaluated herein, expected differences in conformational exploration between the primary (Fig. 3) and negative control (Fig. 4) simulations can be appreciated.

When evaluated by this diagnostic method, the primary simulations show higher RMSD values and less monotonous profiles compared to the negative control simulations, with conformational exploration followed by some degree of return to families of structures similar to the starting structure. The negative control simulations show only gradual drift from the starting structure, and even this only during a prolonged energy equilibration phase lasting up to 400 ps. This is consistent with the dampened motions observed in analogous protein simulations reported by Guenot and Kollman using a distance dependent dielectric [33, 34].

Significance of the results

The major hypothesis of this work, that free MD simulations can generate relevant conformational ensembles which improve agreement with experimental NMR constraint sets, was confirmed for the test case analyzed herein. The 300 ps primary simulations led to significant improvements in agreement with NOE distance constraints for the sheared conformer and improvements in overall agreement with *J* coupling constraints for both conformers, compared to highly refined single static NMR structures. The generality of this result is unknown and can only be ascertained by application to a large variety of test cases.

The result contrasts with a prior study with peptides, involving a much longer (50 ns) simulation, in which agreement with NMR data did not improve at any

point in the simulation [1]. In this study, there was considerable worsening of agreement initially, followed ultimately by equilibration and a stable and modest worsening of agreement compared to the starting structure.

The differences between our result and that of Daura et al. [1] require further investigation. The system studied by Daura et al. had approximately 40 distance constraints, compared to our systems with approximately 200 constraints on a comparable number of monomers. The Daura et al. [1] simulation featured major conformational changes, and transition states between these conformations may have affected the agreement of the ensemble with experiment. Major conformational changes did not occur in our simulation. This may have been due to insufficient simulation length, but it is notable that significant worsening of agreement was observed at the earliest timepoints in the Daura et al. simulation [1]. Alternatively, our system may have less inherent flexibility: this can only be determined by more prolonged simulation. MD simulations explore motions on a variety of timescales, and those on the longest timescales may not have been accurately represented by our simulations or even by a 50 ns simulation. However, if the majority of the NMR constraints in our system were affected primarily by fast motions, overall improvement in *R* factors could have been achieved in spite of failure to represent the slow motions affecting a minority of constraints. The larger number of constraints in our system compared to Daura et al. [1] may have enhanced the ability of the majority of the constraints to dominate the *R* factors. We note that in systems in which the dominant motions affecting agreement with NMR data are slow relative to the timescale of MD simulations, MD simulations would be expected to show limited utility as a structure refinement tool. Finally, due to non-linear averaging, NOEs can be seen between protons which are close to each other only part of the time. This can lead, in highly constrained systems, to NOE constraints which appear contradictory if an attempt is made to infer a single static structure from them. This provides clear opportunities for MD ensembles to improve agreement with experiment, and due to the larger number of experimental constraints in our system compared to that of Daura et al. [1], our simulations may have had more of these opportunities.

The secondary hypothesis of our study was that the ability of free MD simulations to improve agreement would be sensitive to details of the MD methodology, requiring valid simulation techniques. This hypothesis was also confirmed for this test case.

Negative control simulations (500 ps), performed using a method that is no longer current and is expected to dampen molecular motions, did not lead to improved agreement with experiment. While agreement of the ensemble for the sheared conformer was comparable to that for the single static NMR structure, agreement for the imino conformer was significantly worsened by the MD ensemble compared to the single static NMR structure. The failure of the negative control simulations to improve agreement with NMR constraints correlated with reduced exploration of conformational space as evidenced by 2D-RMSD plots, and delayed equilibration as demonstrated by energy plots.

It is of interest that the negative control simulations led to significant worsening of agreement with experiment despite their limited exploration of conformational space. Given that the structures remained close to the starting structures, one might have expected neither improvement nor worsening. However, the worsening was significant for the imino conformer, comparable to that reported by Daura et al. [1] for their simulations in which major conformational changes occurred. This comparison suggests agreement with NMR constraints is indeed sensitive to the presence of subtly erroneous conformers, as likely occurred in our negative control simulations. In the case of Daura et al. [1], maintaining agreement with NMR constraints in the context of large conformational changes is a much more demanding test.

Both our primary and negative control simulations were brief and exploratory and should not be viewed as definitive. It is likely that methodologic enhancements would have led to further improvements in the performance of the primary simulation, further strengthening our conclusions. Moreover, the results suggest that the ability to improve agreement with NMR datasets can be sensitive to the properties of MD methods, raising the possibility that this criterion could be applied to more definitive methods development.

These results may have been altered or improved if longer simulations had been performed. Important conformational changes occurring on time scales of 1–20 ns have been seen in MD simulations of nucleic acids and proteins [1, 2, 14], including nucleic acids with G-A mismatches [47]. It would be of interest to perform longer simulations in this system and follow the *R* factors as a function of time, as done by Daura et al. [1]. This would reveal whether initial changes in *R* factor correlate with changes seen at equilibrium. Achievement of a stable *R* factor may be an indication of the time to reach equilibrium for motions affecting NMR constraints. The time to reach a stable *R* factor

likely varies depending on the inherent flexibility and motions of the system under study. The longest time-scale motions are not currently accessible to MD.

Further insights might be achieved by executing a larger number of simulations from a variety of starting structures or starting velocities [17]. Multiple simulations can give insight into the variability of performance between successive simulations on the same system utilizing the same methods, an issue which is not addressed by this study.

It is also possible that the absence of divalent cations in the simulation may have altered the results. Divalent cations can dramatically affect RNA conformation [48, 49], and high affinity binding sites may be demonstrated by NMR methods [50] or predicted theoretically using Brownian dynamics [51]. Divalent cations bind to Watson–Crick base pairs flanking G-A mismatches in both the DNA/RNA and RNA hammerhead ribozymes, as well as in a novel curved RNA species [17, 18, 52]. The NMR data sets used in this study were generated using RNA that was dialyzed extensively against chelating agents, but this does not fully rule out the possibility of very tightly bound divalent cations. It is notable, however, that the addition of 10 mM MgCl₂ did not alter the imino proton spectrum of the sheared conformer [20]. Neither did the addition of MgCl₂ affect the thermodynamics of tandem mismatch loop closure in a related sequence [26]. MD simulation of divalent cations is difficult due to the long timescale associated with magnesium relocations as well as the effects of charge transfer and major polarization which are not accounted for [53–55].

Some of the observed improvements in agreement with the NMR data may have been realized by time averaging alone. In this regard, it would be of interest to determine if this degree of improved agreement with NMR constraints could be achieved by further refinement of the single optimized NMR structures subject to time averaged restraints [4–8], and whether such refinement leads to the same degree of conformational exploration as judged by 2D-RMSD plots.

A related question is whether comparable enhancements in agreement with NMR data could have been achieved by utilizing the entire ensemble of static NMR structures generated by conventional NMR refinement. This approach could provide additional degrees of freedom and ensure that all structures utilized satisfy the NMR constraints to some degree. In such an approach, one could either weight all static structures from the NMR ensemble equally when generating *R* factors, or allow the weights themselves to vary as independent degrees of freedom in the optimization. In this regard, MD simulation has the

advantage of being a technique which attempts to specify the degree to which various conformational states are occupied, independently of the NMR data, utilizing fundamental principles inherent in force fields and dynamics methodology.

MD ensembles which demonstrate improved agreement with NMR data may be useful for drug design. Conformers in the ensemble may be classified by cluster analysis [1] and ligands designed to bind to representative conformers from different clusters. Conformational substates can also be analyzed in terms of dipole moment, which has previously been shown to contribute to binding equilibria in drug–nucleic acid complexes [16].

In summary, we have demonstrated in this test case that brief free MD simulations can generate conformational ensembles with enhanced agreement with NMR data from unlabeled RNA compared to highly refined single static structures. Comparison with a negative control suggests that the chosen criterion of enhancing agreement with NOEs and coupling constants is sensitive to MD methodologic differences. Thus, free MD may be a useful tool for NMR refinement, and utility in NMR refinement may in turn be a criterion for refining MD methods.

Acknowledgements We would like to thank Wendy Cornell and James Dunbar for advice regarding use of AMBER, Mark Duffield for computer system administration, and Donald Emerson, John SantaLucia, Jr., Douglas Turner, and Eric Westhof for helpful discussions. We are also indebted to Drs. SantaLucia and Turner for sharing their NMR data.

Appendix: Simulation Methods

Force fields

Primary (PME) simulations used the AMBER 4.1 force field [28, 29].

Negative control (DDD) simulations were performed in SYBYL 5.5, 6.1a, and 6.22 in a force field parameterized to be identical to the AMBER 4.1 force field in which the primary (PME) simulations were done [28, 29].

For the sodium counterions, the Lennard-Jones energy parameter ϵ was calculated at 0.0028 kcal/mol and the van der Waals radius at 1.87 Å [56].

Preparation and minimization of conformers

Primary simulation

Structure preparation was performed using standard methods with the AMBER 4.1 [57] programs PREP,

LINK, EDIT, and PARM. Sodium counterions were added to the RNA NMR structures, and the resulting neutral molecule was placed in a box of TIP3P waters [58, 59].

The water molecules were relaxed in SANDER by fixing the RNA and sodium atoms and performing 200 steps of steepest descent minimization. PME [60] was used with the direct sum tolerance set to 0.000001 and the B-spline interpolation order set to 3, with periodic boundary conditions. The periodic box measured $51.9 \times 38.9 \times 38.5$ Å for r(GGCGAGCC)₂ and $51.5 \times 39.7 \times 44.6$ Å for r(GCGGACGC)₂.

SHAKE [61] was applied to all bonds with a tolerance of 0.0005 Å, and the van der Waals cutoff was 9.0 Å. The minimization was performed using an initial step length of 0.1, a maximum step length of 0.5 and a gradient convergence of 0.001.

Negative control simulation

NMR conformers were subjected to simplex optimization and 100 steps of steepest descents minimization with a gradient termination criterion of 0.1 kcal/mol Å, a distance dependent dielectric constant of 1, a non-bonded cutoff of 8 Å, no periodic boundary conditions, and the AMBER 4.1 charges and force field. The structures were nearly identical to the starting structures, hydrogen bonding geometries were maintained, and the structures improved in terms of bond angles and van der Waals energies.

The molecule was then solvated with TIP3P waters using the Sybyl algorithm Silverware [62]. Sodium atoms were then added by replacing the water molecule closest to each phosphate, neutralizing most of the unfavorable electrostatic interactions. The solvated systems with sodium counterions were then subjected to simplex optimization and 100 steps of steepest descents minimization with RNA atoms fixed, termination criterion gradient 0.1 kcal/mol Å, distant dependent dielectric of 1, non-bonded cutoff of 8 Å, AMBER 4.1 force field and charges, and periodic boundary conditions within cubic periodic boxes 36.1 Å on a side. Nine hundred steps of Powell minimization were then undertaken under the same conditions with RNA atoms fixed.

The minimized, solvated conformers with counterions were then further minimized utilizing a low temperature, constant temperature and volume dynamics protocol under the same conditions, with RNA atoms fixed. Five 1 picosecond (ps) intervals were used with set temperatures of 5, 4, 3, 2, and 1 K. The SHAKE algorithm was applied to all H containing bonds. Finally, 100 steps of Powell minimization were performed with RNA atoms fixed.

Dynamics

Primary simulation

Constant volume and temperature (NVT) molecular dynamics was performed using SANDER (AMBER 5.0) with a time step of 2 fs. The Berendsen algorithm for temperature scaling [63] was used. Electrostatics were evaluated with a constant dielectric of 1.0 using PME. SHAKE was applied to all hydrogen-containing bonds. The van der Waals cutoff was 9 Å. Distance constraints were applied to the three hydrogen bonds (G-NH₂:C-O2, G-NH:C-N3, and G-O6:C-NH₂) at each of the terminal GC pairs (G1:C16 and G8:C9) with an equilibrium distance of 1.9 Å and a force constant of 10 kcal/mol.

The water molecules were first allowed to relax in a 5-ps run at 298 K with the RNA and sodium atoms fixed and with a temperature-scaling time constant of 0.40 ps. The sodium atoms were then included in the simulation for successive 1-ps runs at 100, 200, and 300 K, with the same temperature scaling time constant.

Finally, all atoms were included in a series of 1-ps runs at 50, 100, 150, 200, and 250 K and a 5-ps run at 298 K with a temperature scaling time constant of 0.20 ps. Atom velocities were re-generated at each different temperature. The data generation run was for 300 ps at 298 K, with a 0.2 ps temperature scaling time constant and data collected every 25 steps (50 fs).

Negative control simulation

NVT molecular dynamics runs (500 ps) were performed in Sybyl 6.1a (for (rGGCGAGCC)₂) and Sybyl 6.22 (for (rGCGGACGC)₂). The runs consisted of nine 500 fs warming periods at 5, 33, 67, 100, 133, 167, 200, 233, and 267 K, followed by 495.5 ps at 300 K. The initial velocity distribution was Boltzmann for the first period, whereas the previous velocity distribution was carried over for subsequent periods. Integration step size was 1 femtosecond (fs), conformational snapshots were taken every 25 fs, and the temperature coupling constant was 100 fs. The SHAKE algorithm was applied to all H-containing bonds. The force field was that of AMBER 4.1 with a distance dependent dielectric constant of 1, a non-bonded cutoff of 8 Å, and periodic boundary conditions.

References

- Daura X, Antes I, van Gunsteren WF, Thiel W, Mark AE (1999) *Proteins Struct Funct Genet* 36:542
- Reblova K, Spackova N, Sponer JE, Koca J, Sponer J (2003) *Nucl Acids Res* 32:6942
- Clare GM, Gronenborn AM (1989) *Crit Rev Biochem Mol Biol* 24:4479
- Torda AE, Scheek RM, van Gunsteren WF (1989) *Chem Phys Lett* 157:289
- Torda AE, Scheek RM, van Gunsteren WF (1990) *J Mol Biol* 214:223
- Pearlman DA, Kollman PA (1991) *J Mol Biol* 220:429
- Schmitz U, Kumar A, James TL (1992) *J Am Chem Soc* 114:10654
- Schmitz U, Ulyanov NB, Kumar A, James TL (1993) *J Am Chem Soc* 234:373
- Lipari G, Szabo A (1982) *J Am Chem Soc* 104:4546
- Peng JW, Wagner G (1992) *J Magn Reson* 98:308
- Kay LE, Torchia DA, Bax A (1989) *Biochemistry* 28:8972
- Schneider DM, Dellwo MJ, Wand AJ (1992) *Biochemistry* 31:3645
- Akke M, Fiala R, Jiang F, Patel D, Palmer AG III (1997) *RNA* 3:702
- Foloppe N, Nilsson L (2005) *J Phys Chem* 109:9119
- MacKerell AD, Banavali NJ (2000) *J Comput Chem* 21:105
- Gallego J, Ortiz AR, Gago F (1993) *J Med Chem* 36:1548
- Auffinger P, Louise-May S, Westhof E (1995) *J Am Chem Soc* 117:6720
- Beveridge DL, Ravishanker G (1994) *Curr Opin Struct Biol* 4:246
- McConnell KJ, Nirmala R, Young M, Ravishanker G, Beveridge DL (1994) *J Am Chem Soc* 116:4461
- SantaLucia J Jr, Turner DH (1993) *Biochemistry* 32:12612
- Wu M, Turner DH (1996) *Biochemistry* 35:9677
- Jovine L, Hainzl T, Oubridge C, Scott WG, Li J, Sixma TK, Wonacott A, Skarzynski T, Nagai K (2000) *Structure* 8:527
- Pley HW, Flaherty KM, McKay DB (1994) *Nature (London)* 372:68
- Scott WG, Finch JT, Klug A (1995) *Cell* 81:991
- Heus HA, Wijmenga SS, Hoppe H, Hilbers CW (1997) *J Mol Biol* 271:147
- Xia T, McDowell JA, Turner DH (1997) *Biochemistry* 36:12486
- Wöhnert J, Dingley AJ, Stoldt M, Görlach M, Grzesiek S, Brown LR (1999) *Nucl Acids Res* 27:3104
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) *J Am Chem Soc* 117:5179
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA (1996) *J Am Chem Soc* 118:2309
- York DM, Wlodawer A, Pedersen LG, Darden TA (1994) *Proc Natl Acad Sci USA* 91:8715
- Zichi DA (1995) *J Am Chem Soc* 117:2957
- Reyes CM, Kollman PA (1999) *RNA* 5:235
- Guenot J, Kollman PA (1992) *Protein Sci* 1:1185
- Guenot J, Kollman PA (1993) *J Comp Chem* 14:295
- Sklenar V, Miyashiro H, Zon G, Bax A (1986) *FEBS Lett* 208:94
- Lindauer K, Bendic C, Sühnel J (1996) *CABIOS* 12:281
- Withka JM, Swaminathan S, Beveridge DL, Bolton PH (1991) *J Am Chem Soc* 113:5041
- Withka JM, Swaminathan S, Srinivasan J, Beveridge DL, Bolton PH (1992) *Science* 255:597
- Tropp J (1980) *J Chem Phys* 72:6035
- Boelens R, Koning TMG, Kaptein R (1988) *J Mol Struct* 173:229
- Bell RA, Everett JR, Hughes DW, Coddington JM, Alkema P, Hader A, Neilson T (1985) *J Biomol Struct Dyn* 2:693

42. Hilbers CW (1979) In: Shulman RG (ed) Biological applications of magnetic resonance. Academic Press, New York, pp 1–43
43. Robillard GT, Reid BR (1979) In: Shulman RG (ed) Biological applications of magnetic resonance. Academic Press, New York, pp 45–112
44. Wuthrich K (1986) NMR of proteins and nucleic acids. Wiley Interscience, New York, pp 24
45. Sklenar V, Brooks BR, Zon G, Bax A (1987) FEBS Lett 216:249
46. Sklenar V, Feigon J (1990) Nature 345:836
47. Spackova N, Berger J, Sponer J (2000) J Am Chem Soc 122:7564
48. Jossinet F, Paillart JC, Westhof E, Hermann T, Skripkin E, Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (1999) RNA 5:1222
49. Misra VK, Draper DE (2002) J Mol Biol 317:507
50. Hansen MR, Simorre JP, Hanson P, Mokler V, Bellon L, Beigelman L, Pardi A (1999) RNA 5:1099
51. Hermann T, Westhof E (1998) Structure 6:1303
52. Baeyens KJ, De Bondt HL, Pardi A, Holbrook SR (1996) Proc Natl Acad Sci USA 93:12851
53. Reblova K, Spackova N, Stefl R, Csaszar K, Koca J, Leontis NB, Sponer J (2003) Biophys J 84:3564
54. Auffinger P, Bielecki L, Westhof E (2003) Chem Biol 10:551
55. Sponer J, Sabat M, Gorb L Leszczynski J, Lippert B, Hobza P (2000) J Phys Chem B 104:7535
56. Aquist J (1990) J Phys Chem 94:8021
57. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, Ferguson DM, Seibel GL, Chandra Singh U, Weiner PK, Kollman PA (1995) AMBER 4.1. University of California, San Francisco
58. Jorgensen WL (1981) J Am Chem Soc 103:335
59. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
60. Darden TA, York D, Pedersen L (1993) J Chem Phys 98:10089
61. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327
62. © 1990, Rohm and Haas Co. Blanco, M (1991) J. Comput. Chem., 12(2):237
63. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81(8):3684