Published in final edited form as:

J Magn Reson. 2016 April ; 265: 1–9. doi:10.1016/j.jmr.2016.01.006.

Direct NOE Simulation from Long MD Trajectories

G. Chalmers, J. N. Glushka, B. L. Foley, R. J. Woods, and J. H. Prestegard*

Complex Carbohydrate Research Center, University of Georgia, Athens GA, 30602

Abstract

A software package, MD2NOE, is presented which calculates nuclear Overhauser effect (NOE) build-up curves directly from molecular dynamics (MD) trajectories. It differs from traditional approaches in that it calculates correlation functions directly from the trajectory instead of extracting inverse sixth power distance terms as an intermediate step in calculating NOEs. This is particularly important for molecules that sample conformational states on a timescale similar to molecular reorientation. The package is tested on sucrose and results are shown to differ in small but significant ways from those calculated using an inverse sixth power assumption. Results are also compared to experiment and found to be in reasonable agreement despite an expected underestimation of water viscosity by the water model selected.

Graphical abstract

Keywords

Spin relaxation; Simulation; Nuclear Overhauser Effect; Sucrose; Molecular Dynamics

Introduction

Nuclear Overhauser Effects (NOEs), which stem from cross-relaxation between proton pairs in NMR experiments, have provided a great deal of 3D structural information on both large and small molecules in solution [1–3]. When molecules are rigid, the interpretation of NOEs

*Corresponding Author: jpresteg@ccrc.uga.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in terms of 3D structure can be straightforward using the inverse sixth power relationship of the initial build-up rate for the NOE to the inter-proton distance [2]. When the measurement of initial rates is not practical, complete relaxation matrix analysis can account for the effects of multiple relaxation pathways on measurements at longer times and still allow use of this simple distance relationship [4, 5]. When large numbers of NOEs constraining each structural element can be measured, it is even possible to interpret NOEs more qualitatively and still derive a consistent model for a 3D structure [1, 6]. However, when molecules are flexible and sample multiple conformational states, the situation becomes more complex. Appropriate inclusion of the effects of internal motion must then be considered. We present here a software package, based on use of long molecular dynamics (MD) trajectories, to properly simulate NOEs in these cases. Flexible molecules have attracted more interest recently, particularly in studies of the biological roles of intrinsically disordered proteins [7–11], oligomeric carbohydrates [7, 12, 13], and single stranded polynucleotides [14–16]. We anticipate that our simulation package will be useful in properly interpreting NOEs in these cases.

Molecular dynamics simulations provide useful models for predicting the states populated by flexible molecules, as well as their overall tumbling behavior in solution. Traditionally, the use of MD simulations to interpret NOE data has been limited by the relatively short time scales accessible to the simulation. However, with the advent of GPU-based simulation code, it is now possible to extend the simulations into the microsecond regime, enabling the direct computation of key NMR properties. In cases where a limited number of conformational states can be identified by conformational searching or clustering of frames in long MD trajectories it has become common practice to interpret NOEs as depending simply on an average over $1/r^6$, where r is the distance between spin pairs in various conformational states [17, 18]. However, this type of averaging is only appropriate when the distance and angular variations leading to spin relaxation can be assumed to be uncorrelated, as they are when the rates of transition between states are slow compared to molecular tumbling [19]. Formal treatments that avoid this assumption have been known for some time [20, 21], and approaches that use MD data to directly calculate spin relaxation parameters have been demonstrated [22, 23]. One of these approaches was, in fact, motivated by concern over the limitations of $1/r^6$ or $1/r^3$ averaging in analysis of NOE (or ROE) data on flexible peptides [23]. These methods calculate correlation functions directly from a trajectory, rather than extracting an average of $1/r^6$ (or $1/r^3$), as a step in calculating the relaxation parameters. Use of such approaches for NOE interpretation is, however, still rare, in part because of the need for long MD simulations, and in part because of the lack of convenient tools for comparing MD-derived values to experimental NOEs. We present a version of such tools here.

The tools we present here both calculate NOEs from an MD trajectory, and provide a means of comparing the calculated NOEs to experiment. The general process takes as input an MD trajectory (long enough to fully sample the conformational states), assesses its suitability for representation of motional properties, calculates correlation functions for pair-wise dipolar interactions directly from the trajectory, and enters resultant relaxation parameters into a complete relaxation matrix approach [24]. This process is integrated into a software package (MD2NOE) that is being made available to the scientific community.

The utility of the tool is illustrated with analysis of NOE data acquired for sucrose in aqueous solution. Sucrose (β -D-fructofuranosyl-(2-1)- α -D-glucopyranoside) is a disaccharide that exhibits a range of internal motions, both slow and fast compared to its overall tumbling time; sucrose has become a standard for testing the validity of MD simulations, as well as analyzing NMR spin relaxation measurements [25–28].

Theory

Dipole-dipole spin relaxation is dependent on fluctuations in interactions between pairs of nuclei that occur at frequencies that match the energies of transitions allowed by those interactions. The time dependent nature of those fluctuations is conveniently embodied in the well-known correlation function given in Equation 1 [3]. Here dd is the dipolar interaction constant at unit distance, $r(t)$ is the distance between the pair of nuclei in question, Y_2^0 are spherical harmonics dependent on angles Ω of inter-nuclear vectors relative to the laboratory field, t is time, τ is a separation between two time points, and the brackets indicate averaging over time (in our case the length of an MD trajectory).

$$C(\tau) = \langle (dd)^2 \frac{1}{r^3(t)} \times Y_2^0(t) \times \frac{1}{r^3(t+\tau)} \times Y_2^0(t+\tau) \rangle_t \quad (1)$$

The two $1/r^3$ terms can be factored out and averaged separately only if angular and distance variations are uncorrelated. This can be done in the two limits where motions are either fast compared to molecular tumbling or slow compared to molecular tumbling. The two limits clearly give different results, the square of the average of $1/r^3$ multiplied by internal angular variations in the former case, and the average of $1/r^6$ in the latter. An extreme example is the case where fast internal motions allow one or both nuclei to sample a spherical distribution, in which case the angular and distance variations cancel exactly, and the effective distance is that between the centers of the spheres [29]. In that case the presence of substantial internal motion is completely obscured. At the other extreme, the $1/r^6$ averaging of distances that are sampled in slowly-interconverting conformational states heavily weights conformers having the closest approach of spin pairs, leading to an apparent increase in the population of close approach conformers. Alternatively, snapshots extracted at regular time points from a dynamics trajectory can be used to calculate the correlation function in Equation 1 directly, provided the trajectory is sufficiently long to sample a uniform distribution of molecular orientations, as well as all changes in conformation capable of contributing to the average NMR data. For sucrose, the molecule we use for illustration here, this amounts to simulations spanning at least a hundred nanoseconds.

The power distribution across frequencies that give rise to various spin relaxation processes is given by the real-values cosine Fourier transform of the correlation function in Equation 1. Homonuclear ^1H - ^1H NOEs, measured from the initial buildup of the resonance of one proton after the saturation, inversion, or periodic perturbation of the spin state populations of another proton, are the result of competition between zero and two quantum transitions occurring at the difference and sum of single proton transition frequencies. The rate constants corresponding to this process are called cross-relaxation rates, and are given the

symbol σ . These rates are given in equation 2. Here, J_{ij} is the Fourier transform of the correlation function (i.e. spectral density function) for a pair of protons (I,j), in our case including both distance and angle variations. It is sampled at the relevant frequencies, 0 and 2ω , where ω is the proton Larmor precession frequency in radians/s. Hence, σ can be evaluated for any proton pair using the correlation functions described in Equation 1.

$$\sigma_{i,j} = \frac{(dd)^2}{4} \{-J_{ij}(0) + 6J_{ij}(2\omega)\} \quad (2)$$

Experimentally, extraction of an initial slope from an NOE build-up curve is challenging because of the presence of indirect transfers and processes causing relaxation back to equilibrium. The rate constant for the latter relaxation process, ρ , is given in Equation 3 where the summation is over all spins having a significant interaction with spin i (typically within a distance of 4 or 5 Å).

$$\rho_i = \frac{(dd)^2}{4} \sum_{j \neq i} \{J_{ij}(0) + 3J_{ij}(\omega) + 6J_{ij}(2\omega)\} \quad (3)$$

In an NOE experiment, the ratio of signal to noise at sufficiently short times for a direct measure of σ is frequently not viable. However, the system of interacting protons giving rise to effects at longer times can be modeled in a formalism called complete relaxation matrix analysis [5, 24]; it can be used to produce NOE build-up curves that may be compared to experiment. Here, we have implemented a complete relaxation matrix simulation using σ and ρ terms calculated from MD trajectories. We should point out that despite the designation of complete relaxation matrix, most implementations, including ours, neglect interference from cross correlation effects that occur in the presence of multiple dipolar interactions. In principle, these interference effects can be integrated into our approach by exploring an expanded set of spin-interaction. We leave this for future work.

Software package

The software package (MD2NOE) described here is a part of a suite of programs that evaluate MD trajectories, relate them to structural parameters, and simulate other experimental NMR observables including spin-lattice relaxation rates, spin-spin relaxation rates and $^3J_{HH}$ scalar couplings. The programs are written in C++, and currently run via command line under LINUX operating systems. There are options in the programs to specify the formatting of the figures (e.g. jpeg, eps, eps-latex, pdf, or no figures). The programs, as they are created and tested, are made available at glycam.org/nmr. The programs associated with this publication will be available at glycam.org/publication-materials/md2nmr.

MD2NOE takes as input an MD trajectory from AMBER, and the associated topology file (Figure 1, Module 1). Before entering the main program the trajectories should be checked for problems, such as inadequate sampling of conformational states or departures from isotropic sampling of orientation during the simulation. This may be done with auxiliary

programs represented in Figure 1 as Module 2. One of these auxiliary programs, “TRAJECTORY”, uses the trajectory to generate plots and text files showing inter nuclear distances for selected pairs of protons as a function of time. Visual examination of these trajectory representations can assure that steady-state behavior has been reached for the parts of the trajectory to be used in the calculation, and can identify transitions between conformers that occur frequently enough in the trajectory to allow accurate representation in the correlation function (several tens of transitions is desirable).

The second auxiliary program, “ISOTROPY_CHECK” checks for adequate isotropic sampling of rotational states. One factor that can cause non-isotropic behavior is the choice of solvent box dimensions. That is, the ‘box’ containing the solute and any explicitly modeled solvent (water in this work) should be sufficiently large to allow the solute to tumble freely without being affected by its neighboring images, and it should be cubic so that only isotropic position scaling can be employed to control atomic velocities. Non-isotropic position scaling, which is a standard option in AMBER, can lead to non-uniform sampling of molecular orientations and will introduce artefactual differences in simulated cross relaxation rates for proton pairs.

After selecting a trajectory, correlation functions are calculated according to Equation 1 as a frame-by-frame average over the length of the trajectory, and plotted for visual examination unless the option of no figures is used (Module 3, Figure 1). The numerical information is also saved in text files. When trajectories are sampled adequately, and represent statistically random processes, such as rotational Brownian motion, the correlation functions are sums of smooth exponential decay terms that approach zero at long times. Departures from this behavior usually indicate trajectories that are too short to fully sample the relevant molecular motions. It is also useful to compare correlation function decay times, for proton pairs in a rigid part of the molecule, to correlation times predicted by the Stokes formula, using an appropriate hydrodynamic radius, temperature, and solvent viscosity, to see if overall rotational timescales are properly represented.

In the next step (Module 4, Figure 1) a numerical Fourier transform of the correlation function is performed and relaxation parameters, ρ and σ , are calculated according to Equation(s) 2 and 3. The behavior of longitudinal magnetization associated with any spin is typically described in terms of a set of differential equations (Solomon equations) that use ρ and σ values as coefficients. This set of equation can be written in matrix form in which ρ and σ values are diagonal and off diagonal elements of a matrix often called a complete relaxation matrix [24, 30, 31]. Output of Module 4 includes a representation of the relaxation matrix in tabular form.

The program then (Module 5, Figure 1) generates NOE build-up curves based on the type of experiment to be simulated (selective inversion, or selective saturation). Diagonalization and exponentiation of the relaxation matrix provides solutions for the time evolution of an initial set of magnetization values that is specific for each of these experiments. The NOE curve information is saved in text files and figures are also given unless the option of no figures is used.

A potential source of deviation from experiment is inaccurate conformer populations from the trajectory. Prediction of accurate populations of particular conformational states requires a very high-quality force field, as well as adequate sampling of all conformers. For this reason, there is a module (Module 6, Figure 1) that allows populations of selected states to be determined that lead to a best fit to the experimental data by deleting or inserting segments of the trajectory that represent unique states. NOEs can then be recalculated iteratively to reach agreement with experiment. This latter approach has the benefit of leveraging the power of MD simulations to identify relevant conformational states, without requiring the trajectory to be long enough to accurately predict relative populations of the states.

Most information is entered as command line arguments. Since there are a large number of options, these are most easily handled by editing a control script (bash) and calling this script to execute the program. An example of a script with argument definitions and a call to the program, MD2NOE, follows:

```
#!/bin/bash
# command line arguments for the program "MD2NOE" are defined as:
# argv[1] topology file name
# argv[2] MD trajectory file name
# argv[3] the total time of the simulation in ns
# argv[4] the display time for the correlation function in ns
# argv[5] the sampling interval for the trajectory file
# argv[6] the magnetic field in Tesla ( $B_0$ )
# argv[7] percentage of the correlation function Fourier transformed
# argv[8] total time in ms of the NOE curve
# argv[9] interval between NOE points calculated (ms)
# argv[10] cutoff distance for proton pairs included in the calculation
# argv[11] format for plots (jpeg,eps,pdf)
# argv[12] designates which NOEs are output(all/notall - reads from file)
# argv[13] name of directory for output
# argv[14] to argv[14-1+nprotons] equilibrium intensities for each proton
# argv[14-1+nprotons] to argv[14-1+2*nprotons] initial state intensities
# call to the program md2noe with example input for sucrose. Of the 22
# protons in sucrose all but the fourth proton in the topology file are
# inverted in the initial state
./MD2NOE sucrose_noWAT_tip3.prmtop prod.crd 100 10 1 14.0 .02 5000 10 5
jpeg all sucrose_noes 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 -1 -1 -1
1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
```

The program outputs text files and figures characterizing all spin pairs. To facilitate examination of the potentially large number of output files, separate programs have been coded to extract typically-desired information. For example the program,

“NOE_EXTRACT”, finds the NOE for a particular spin pair at a particular mixing time; it too can be called by a bash script:

```
#!/bin/bash
# command line arguments for the program "NOE_EXTRACT" are defined as:
# argv[1] directory for input
# argv[2] directory for output
# define noe of interest
# argv[3] residue 1
# argv[4] atom 1
# argv[5] residue 2
# argv[6] atom
# define pair to be used as an NOE reference
# argv[7] residue 1
# argv[8] atom 1
# argv[9] residue 2
# argv[10] atom 2
# argv[11] mixing time
# the following is an example call to the program
./NOE_EXTRACT test_in test_out 2CU H11 2CU H12 2CU H61 2CU H62 100
```

METHODOLOGY

Trajectories

MD trajectories of sucrose in water were generated using the pmemd.MPI utility in AMBER 12 [32]. The GLYCAM06 force field (version Glycam06h) [33] was used for the solute, and the TIP3P[34] solvent model was used for the water. The GLYCAM06 force field has been validated previously by comparing simulated with experimental structures. However, additional validation for sucrose has been obtained for this study by comparing calculated and experimental scalar proton-proton J couplings. Water models can be evaluated based on their ability to reproduce a number of solvent properties. The TIP3P water model, for example, is known to underestimate self-diffusion constants, by a factor of 2.2 at 25°C and 1 atm. [35]. However, we expect these errors to result in a predictable scaling of many NOE effects and have chosen TIP3P in the interest of computational efficiency. If desired, a more complex water model, TIP5P, for example, can be used.

The initial model of sucrose immersed in a cubic box of water molecules was built using the tleap utility of AMBER12 [32]. The sucrose was constructed from template structures for β -D-fructofuranose (GLYCAM residue template 2CU [36]) and α -D-glucopyranose (residue template 0GA). The solvent box was built up from repeats of a small cubic template of pre-equilibrated water molecules (template name TIP3PBOX) so that the minimum distance from the sucrose to any box wall was 12 Å and that, overall, the box, containing ~4000 water molecules, was approximately cubic. This model was then subjected to the following simulations steps, using the pmemd.MPI module. First, the positions of all atoms in the

system were energy-minimized for 1000 steps using steepest descent under non-periodic conditions. For the following steps, periodic boundary conditions were employed, the lengths of bonds including hydrogens were restrained using the SHAKE algorithm in the manner recommended for use with TIP3P, the non-bonded cutoff was set to 8 Å, electrostatics in the periodic system were handled by particle mesh Ewald, and the time step was 2 fs. After minimization, the system was heated at constant volume from 5 K to 300 K in 20,000 steps using the NMR restraints functionality (which is not, in this case, performing any NMR-related function), and the Berendsen thermostat. The system was then switched to constant pressure (isotropic scaling at 1.0 bar) and temperature (Langevin dynamics at 300 K). It was allowed to equilibrate in these conditions for 1,000,000 steps (2 ns). Solute coordinate data was saved every 1000 steps (2 ps) from the following 50,000,000 simulation steps (100 ns). Ten such simulations were run, beginning each simulation with an independent minimization to assure sampling of independent configuration spaces. Data from all ten simulations, a total of 1 μ s of simulation data, were used for this study. Further details regarding the methods and utilities used to generate the trajectories can be found in the AMBER 12 Reference Manual and the AmberTools12 Reference Manual [29], and specific details of the simulation can be obtained by consulting the simulation input and output files available for download at glycam.org/publication-materials/md2nmr.

Experimental NOEs

Data were collected on a 42mM solution of sucrose in D₂O at 25°C on an Agilent 600 MHz Direct Drive spectrometer equipped with a 5 mm cryogenically cooled probe. A standard Agilent pulse program, “NOESY1D”, was used to acquire one-dimensional transient NOE spectra. In this sequence, signals are selected by a pair of Q3 Gaussian cascade pulses flanked by gradients. The particular experiment mimics the generation of a cross peak in a 2D NOE experiment by setting the initial magnetization for the selected pulse to zero. The selective pulse widths varied between 74 and 174 ms depending on the signal being irradiated. A series of mixing times were collected with 512 scans per spectrum to achieve high a signal-to-noise ratio. Data were processed with VNMRJ (Agilent, Inc.) or with MNova (Mestrelab Research) software. Typically, data were multiplied by an exponential weighting function that added 1 Hz to the line width, zero filled to 64K points, Fourier transformed, phase corrected and baseline corrected. Peaks were integrated in MNova using manual selection of integral regions.

RESULTS

Figure 2 shows sucrose in three representative conformations, along with a representation of the trajectory input for examination in Module 2 (Figure 1). The distance between the anomeric proton (H1) on the glucopyranose residue (g) and one of the trans-glycosidic protons (H12) on the C1 carbon of fructofuranose residue (f) is depicted. Variations in this distance are expected to capture transitions between conformers that differ in glycosidic torsional angles as well as the exocyclic angle about the fC1-fC2 bond, ω . In the trajectory representation, the distances have been extracted with snapshots collected at 200 ps intervals from a 100 ns simulation, beginning after the minimization and equilibration. It is clear that there is a broad distribution of conformers with a dominant set having distances between 3.4

and 3.7 Å, and a minor set with a broader distribution of distances between 2 and 3 Å. A distance of 3.5 Å corresponds to a conformer with ϕ and ψ glycosidic angles near 108° and -45°, and with the exocyclic torsion angle, ω , near -60°. The glycosidic angles are similar to those observed in the sucrose crystal structure (108, -55) [37], and structures near this state are predicted to be highly populated in solution [21, 27]. Distances between 2 and 3 Å are also consistent with this glycosidic linkage orientation, but with an ω angle of approximately 180°, or with a second highly populated conformer predicted in previous MD studies with ϕ , ψ , and ω angles near 70, -75, and 180° [28]. Except for very occasional sampling of a much longer distance, transitions between the major states are frequent (occurring on a timescale of a ns). The trajectory is sufficiently long to capture these transitions, and it is also clear that many of the variations in distance occur on a timescale (ns) that is very similar to that of overall tumbling, estimated using Stoke's law for rotational diffusion for a spherical molecule with the molecular weight of sucrose in a dilute D2O aqueous environment at 298 K to be 0.15 ns. This, in turn, is very close to an experimental determination based on ^{13}C spin relaxation measurements [27, 38]. Hence, we would also expect sucrose to be a molecule where the calculation of NOEs from a correlation function based on an MD trajectory would be more appropriate than $1/r^6$ averaging over the simulated distances.

To properly represent spin relaxation phenomena, such as NOEs, the trajectory must fully capture molecular tumbling. Violations of this condition can be detected by examining the projection of the vector connecting any pair of protons whose interaction is minimally affected by internal motion. Using one of the tools, "ISOTROPY_CHECK" of Module 2, Figure 3 shows the average of the x-projection of the vector between gH1 and fH12, as averaging is extended from a few ns to 100 ns, for a simulation performed in a cubic box (12 Å buffer). The average projection starts near 2 Å and decays quickly on a timescale consistent with molecular tumbling. Within one nanosecond it converges to a value close to zero, as expected for isotropic sampling of orientations. A trajectory using the default rectangular (non-cubic) box with sides 10 Å outside the maximum extent of the sucrose molecule in each dimension led to tumbling that did not achieve isotropic orientational sampling as indicated by the averages of the vector projection deviating from zero by more than 20%. This could lead to significant errors in relaxation parameter simulation.

The correlation function for the trans-glycosidic proton pair, gH1 – fH12, calculated using tools of Module 3, is shown in Figure 4. The entire trajectory is 100ns long. τ values separating points to be averaged according to Equation 1 range from zero to 2ns, where the correlation function approaches zero. The correlation function decays in a multi-exponential fashion. Fitting to two exponentials and a small constant offset suggested correlation times of 0.097 ns of amplitude 0.30 and 0.018 ns of amplitude 0.60. The longer correlation time of (0.097 ns) is near the 0.15 ns expected using the Stokes formula for a molecule the size of sucrose in a dilute water solution at 298 K. This does not show quite the factor of 2.2 deviation expected using a TIP3P water model, but correlation times for other pairs, gH1-gH2 for example, are shorter (0.064 ns). Some of this variation can be due to anisotropic motion when interaction vectors between spin pairs point along different rotational axes; correlation times based on hydrodynamic calculations of bead models for sucrose, which take anisotropy of motion into account, show 20–35% variation in correlation times among

diffusional axes [39]. The variation may also come from the limitations of using just a two exponentials to fit the multi-exponential decay of gH1-fH11/12 data; the gH1-gH2 data, in contrast, fit well to a single exponential.

The shorter correlation time (0.018 ns) for the gH1-fH11/12 fit likely represents an effective timescale for internal motions of this trans-glycosidic pair of protons. The magnitude of the function in Equation 1 at $\tau=0$ is a measure of the mean square of the dipole-dipole interaction, but for presentation purposes, values at all times have been normalized with respect to the initial value. The need to employ a constant offset in the fit may be associated with inadequate sampling of an isotropic orientational distribution due either to the limited box size or the length of the trajectory. However, this offset is small (less than 10%), and should not have a significant effect on predicted NOEs.

In Table 1 relaxation parameters, as calculated in Module 4, are presented for a set of proton pairs that includes the gH1-fH12 pair considered in the previous illustrations. This table represents a piece of the complete relaxation matrix that will be used in calculation of an NOE. In the actual matrix the off-diagonal elements would be the cross relaxation rates, σ , for each pair. The diagonal elements would be the total longitudinal relaxation rates, ρ , which include effects from all protons interacting with a particular proton, not just those in the table.

NOE buildup curves are generated in Module 5 using the relaxation matrix discussed above, together with the initial and equilibrium representations of resonance intensities. Curves illustrating the departure from equilibrium intensity for the fH12 resonance and the fH11 resonance, after initial zeroing of the resonance for gH1, are presented in Figure. 5 along with a curve for their sum. The initial and equilibrium conditions were set in the simulation by vectors of intensities. For the case at hand these would be 1,1,-0,...,1,1 and 1,1,1,...,1,1 in which the third entry represents the gH1 proton. The intensity of both the fH11 and fH12 signals clearly rise from a zero departure from equilibrium to a value indicating a positive enhancement of about 1.5% before beginning a decay back toward zero. The initial slopes give the cross relaxation rate for the gH1-fH12 and gH1-fH11 pairs. The positive slope is typical of expectations for a small molecule with correlation times in the extreme narrowing limit. The buildup never reaches expectations for steady state saturation of gH1, because of a continual loss of inverted gH1 magnetization after the initial inversion, indirect transfer to other proton pairs by the fH12 and fH11 protons, and other losses to the lattice as fH12 and fH11 recover to equilibrium.

Also included in Figure 5 is a depiction of experimental data for comparison to simulated curves. The resonances for fH11 and fH12 are degenerate, hence, the sum of the two simulated curves for fH11 and fH12 provides a better comparison to experiment. The agreement is quite good with the simulated initial slope being slightly smaller. Again, we do not see the factor of 2.2 underestimate that we might have expected had cross-relaxation been dominated by molecular tumbling contributions and viscosity had been underrepresented by the TIP3P water model [35]. However, additional contributions come from internal motions in this case, and it is less obvious how viscosity would affect these motions. The more rapid ρ decay of the experimental data at longer times could be due to

contributions not included in our simulation model, such as inter-molecular effects and those from dissolved oxygen in the sample. While both contributions are likely to be small, sucrose molecules are known to weakly associate at high concentrations [40], and the effects of paramagnetic dissolved oxygen are well documented [41].

A more extensive comparison of predicted and experimental NOE data is given in columns 2 and 3 of Table 2 for protons having well resolved resonances. The initial slopes in the table come from the initial slopes of build-up curves such as those in Figure 5 in the case of experiment, or from calculated cross relaxation rates such as those in Table 2 in the case of simulated data. Simulated and experimental initial slopes are again very similar (within 25%) except for the gH1-fH4 pair where the initial experimental slope is a factor of two larger. However, the fact that the initial slope for the gH1-gH2 pair is actually larger than the experimental slope (0.075 vs 0.061 s^{-1}) is somewhat disturbing. This pair is within a glucose ring that should have relatively small conformational fluctuations letting overall tumbling dominate relaxation; also, the correlation function for this pair has a single exponential decay time that is nearly a factor of 2.2 shorter than expectation. The explanation for the similarity of initial slopes is relatively simple; while a correlation time of 0.064 ns is comfortably within the extreme narrowing limit, a correlation time of 0.15 ns is not. A calculation using the gH1-gH2 distance ($2.4 \text{ \AA}$), the correlation time of 0.15 ns , and the analytical formula for the spectral density function of a single exponential decay, one predicts a cross relaxation rate of 0.087 s^{-1} , something that is very close to our simulated value of 0.075 s^{-1} . Moreover, the values for ρ contributions rise with the increase in correlation time by about 60%, an additional factor contributing to the enhanced long time decay of the experimental data presented in Figure 5. Hence, most of our concern relating to experimental vs simulated initial slopes in Table 2 is dispelled by a proper consideration of the effect of a longer correlation time on these slopes [27].

A common practice in the literature is to convert slopes to a distance using the slope of a reference pair and the $1/r^6$ approximation. While this is valid only for a rigid system, it would eliminate some of the error associated with imperfections such as the viscosity representation of the water model. We have, therefore, used the distance between a pair of protons in the conformationally less flexible glucose ring (${}^4\text{C}_1$ chair), namely the gH1-gH2 pair as a reference ($2.4 \text{ \AA}$). Distances calculated in this way for experimental and simulated data are reported in columns 4 and 5 of Table 2, respectively. All agreements between simulation and experimental prediction appear to improve, including the gH1-fH4 pair.

DISCUSSION

The software modules presented clearly provide a viable means of selecting MD trajectories that are suitable for simulation of spin relaxation processes and predicting spin relaxation properties, such as NOEs, from those trajectories. As shown in Table 2, agreement between distances extracted from experimental and simulated initial slopes of NOE build-up curves, using a simple rigid model, seems reasonable with differences of 6% for the average of the fH11-gH1 and fH12-gH1 distances, and 16% for the fH4-gH1 distance. But, both levels of agreement are well outside of experimental error. The rigid model is clearly not correct for these trans-glycosidic pairs, and would not be expected to be correct based on distance

variation plots such as that presented in Figure 2. The primary purpose of our calculating NOEs from an MD trajectory is, in fact, to explore ways of properly adding the effects of motion to NOE interpretation.

A common means of interpreting NOEs in the presence of motion is to assume NOEs are proportional to the $1/r^6$ average over inter-proton distances observed in all frames of the dynamics trajectory. We have included distances calculated from the inverse sixth root of this average in column 7 of Table 2, as well as a simple distance average in column 6 of Table 2. For the gH1:gH2 pair both means of calculating a distance give nearly identical answers and agree well with distances expected for low energy conformers of this molecule. This would be expected for a pair within the glucose ring if there were relatively little internal motion. For the other pairs the results are different, and the $1/r^6$ average produces better agreement with experiment. A question remains, however, as to whether this means of relating an NOE to an MD trajectory is adequate, and whether the calculation of NOEs from correlation functions as described in Equation 1 gives significantly different values.

To address this question we compared the distances calculated from the initial slopes of our simulated NOEs (column 5 of table 2) to the distance from the $1/r^6$ average. If $1/r^6$ averaging were rigorously correct, the distances in columns 5 and 7 would be identical, however, the values for the gH1-fH11 pair and the gH1-fH12 pair, in particular, deviate by 8% and 9% respectively. This suggests that $1/r^6$ averaging is not strictly correct for the large amplitude, rapid internal motions, seen for these pairs in sucrose. Although the actual deviations in distances are not large, there are contexts in which they may be significant. For example, if one wanted to parameterize a force field based on agreement with NOE data, small variations in distance may be important; changes in close contact distances can correspond to several kcal in interaction energy. Another issue arises in the evaluation of populations of different conformers seen in a trajectory. When transitions between these states are slow compared to fluctuations that directly affect relaxation, the observed NOE is the population weighted average of the NOEs in the individual states. The error made in improperly calculating an NOE would then have about six times the effect on the population that it has on an averaged distance; 48–54% for the cases cited above. We should note that errors in relating NOEs to conformational distributions are not limited to molecules carbohydrates like sucrose. A very similar observation was made for flexible peptides on comparing experimental ROE data to simulations which calculated a correlation function directly from an MD trajectory [23].

Knowing that we have a more accurate way to calculate NOEs in the presence of motion, and that our simulations are likely to improve with use of a better water model it is appropriate to consider ways any remaining deviations might be used to improve MD calculations. There are two ways that we might adjust MD based simulations to match experiment. One is to adjust force field parameters. However, calculated energies must be highly accurate to properly simulate populations of distinct conformers. A difference in population of 30% at 298 K corresponds to an energy difference of just 0.65 kJ or 0.16 kcal/mol. This level of fine-tuning would be difficult to implement in, for example, the generation of a Class I MD force field, meant to treat many different molecules with a minimum of molecule-specific parameters. An alternative is to assume that the structures of the

conformational states are properly represented, and simply adjust the time spent in a conformational state until observable NOEs are reproduced. This ability has been introduced in Module 6 (Figure 1). If significant periods of time in the trajectory can be identified, during which a molecule samples only one conformational state, this period of the trajectory can be expanded or truncated by repeating or omitting frames, as long as the frames of the adjusted segment and the frames that follow that segment, are transformed to minimize molecular reorientation at the junctions. After comparison of a simulated NOE to experiment, the process represented in Figure 1 can then be iterated, increasing or decreasing the representation of particular conformers, until simulated NOEs reproduce experimental data. This latter module has not been tested on sucrose because there are few segments in the trajectory where a single set of conformers is populated for more than a few ns. There are, however, cases in other oligosaccharides where residence times are sufficient to apply this module, and an illustration will appear in a subsequent publication.

There are still some fundamental limitations in making MD trajectories match NMR spin relaxation data. The force fields that dictate motional behavior have been derived to simulate conformations and energetics near minimum energies in a complex conformational surface, and do not necessarily accurately reproduce transition barriers between minima. This will clearly affect an ability to simulate the timescales of some motional processes. We don't believe that this is an important issue for sucrose because the motions we see are largely oscillations about local minima with a few transitions over low barriers. However, as we examine more complex systems this is an issue that may deserve some attention. The approach is also dependent on an ability to run sufficiently long trajectories to achieve statistically valid sampling of both molecular reorientation and internal motion. For sucrose this is clearly feasible and given that the power of computational resources is continually improving, the simulation of larger systems will be possible. For even larger systems a modified approach is possible, in which the internal and overall motions are separated, a correlation function for internal motion is calculated using a trajectory that is nevertheless too short to sample tumbling of a large molecule, and these data are combined with a correlation function representing rotational diffusion. Such a process has been described in the literature [42] and is being implemented as an option in our simulation software. Also, modifications to address other limitations, such as the effects of correlated dipolar interactions on relaxation in multi-spin systems are being considered.

Simulations for sucrose incorporating the improvements described above, especially the use of a better water model, are very likely to allow a detailed interrogation of conformational states and motional transitions between those states. Despite the long history of data collection on sucrose, there is still debate as to the level of internal motion about the glycosidic bonds versus exocyclic hydroxymethylenes [26, 27]. We certainly see both types of motion in the trajectory studied. Whether the proportions change with water model and whether the modified trajectories can be validated with fits to a wider range of experimental data remains to be seen.

Acknowledgments

This work was supported by grant from the National Institute of General Medical Sciences of the National Institutes of Health, GM103390, GM033225 and GM100058.

References

1. Wuthrich, K. NMR of Proteins and Nucleic Acids. Wiley; New York: 1986.
2. Neuhaus D, Williamson MP. THE NUCLEAR OVERHAUSER EFFECT IN STRUCTURAL AND CONFORMATIONAL ANALYSIS. 1989
3. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG.; Rance, M.; Skelton, NJ. Protein NMR Spectroscopy: Principles and Practice. 2. 2007.
4. Voegeli B. The nuclear Overhauser effect from a quantitative perspective. Progress in Nuclear Magnetic Resonance Spectroscopy. 2014; 78:1–46. [PubMed: 24534087]
5. Borgia BA, Gochin M, Kerwood DJ, James TL. RELAXATION MATRIX ANALYSIS OF 2D-NMR DATA. Progress in Nuclear Magnetic Resonance Spectroscopy. 1990; 22:83–100.
6. Kosol S, Contreras-Martos S, Cedeno C, Tompa P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. Molecules. 2013; 18:10802–10828. [PubMed: 24008243]
7. Almond A, Bunkenborg J, Franch T, Gotfredsen CH, Duus JO. Comparison of aqueous molecular dynamics with NMR relaxation and residual dipolar couplings favors internal motion in a mannose oligosaccharide. Journal of the American Chemical Society. 2001; 123:4792–4802. [PubMed: 11457289]
8. Eliezer D. Biophysical characterization of intrinsically disordered proteins. Current Opinion in Structural Biology. 2009; 19:23–30. [PubMed: 19162471]
9. Hosek T, Gil-Caballero S, Pierattelli R, Brutscher B, Felli IC. Longitudinal relaxation properties of H-1(N) and H-1(alpha) determined by direct-detected C-13 NMR experiments to study intrinsically disordered proteins (IDPs). Journal of Magnetic Resonance. 2015; 254:19–26. [PubMed: 25771525]
10. Jensen MR, Zweckstetter M, Huang, Backledge M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. Chemical Reviews. 2014; 114:6632–6660. [PubMed: 24725176]
11. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nature Reviews Molecular Cell Biology. 2015; 16:18–29. [PubMed: 25531225]
12. Battistel MD, Azurmendi HF, Yu B, Freedberg DI. NMR of glycans: Shedding new light on old problems. Progress in Nuclear Magnetic Resonance Spectroscopy. 2014; 79:48–68. [PubMed: 24815364]
13. Canales A, Jimenez-Barbero J, Martin-Pastor M. Review: use of residual dipolar couplings to determine the structure of carbohydrates. Magnetic Resonance in Chemistry. 2012; 50:S80–S85. [PubMed: 23280664]
14. Taqi MM, Warmlander SKTS, Yamskova O, Madani F, Bazov I, Luo J, Zubarev R, Verbeek D, Graslund A, Bakalkin G. Conformation Effects of CpG Methylation on Single-Stranded DNA Oligonucleotides: Analysis of the Opioid Peptide Dynorphin-Coding Sequences. Plos One. 2012; 7
15. Eichhorn CD, Al-Hashimi HM. Structural dynamics of a single-stranded RNA helix junction using NMR. Rna-a Publication of the Rna Society. 2014; 20:782–791.
16. Eichhorn CD, Feng J, Suddala KC, Walter NG, Brooks CL III, Al-Hashimi HM. Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. Nucleic Acids Research. 2012; 40:1345–1355. [PubMed: 22009676]
17. Forster MJ, Mulloy B. RATIONALIZING NUCLEAR OVERHAUSER EFFECT DATA FOR COMPOUNDS ADOPTING MULTIPLE-SOLUTION CONFORMATIONS. Journal of Computational Chemistry. 1994; 15:155–161.
18. Blundell CD, Packer MJ, Almond A. Quantification of free ligand conformational preferences by NMR and their relationship to the bioactive conformation. Bioorganic & Medicinal Chemistry. 2013; 21:4976–4987. [PubMed: 23886813]

19. Mulloy B, Forster MJ. Conformation and dynamics of heparin and heparan sulfate. *Glycobiology*. 2000; 10:1147–1156. [PubMed: 11087707]
20. Bruschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR. INFLUENCE OF RAPID INTRAMOLECULAR MOTION ON NMR CROSS-RELAXATION RATES - A MOLECULAR-DYNAMICS STUDY OF ANTAMANIDE IN SOLUTION. *Journal of the American Chemical Society*. 1992; 114:2289–2302.
21. Case DA. Molecular dynamics and NMR spin relaxation in proteins. *Accounts of Chemical Research*. 2002; 35:325–331. [PubMed: 12069616]
22. Prompers JJ, Bruschweiler R. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *Journal of the American Chemical Society*. 2002; 124:4522–4534. [PubMed: 11960483]
23. Peter C, Daura X, van Gunsteren WF. Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *Journal of Biomolecular Nmr*. 2001; 20:297–310. [PubMed: 11563554]
24. Keepers JW, James TL. A THEORETICAL-STUDY OF DISTANCE DETERMINATIONS FROM NMR - TWO-DIMENSIONAL NUCLEAR OVERHAUSER EFFECT SPECTRA. *Journal of Magnetic Resonance*. 1984; 57:404–426.
25. Xia J, Case DA. Sucrose in aqueous solution revisited, Part 2: Adaptively biased molecular dynamics simulations and computational analysis of NMR relaxation. *Biopolymers*. 2012; 97:289–302. [PubMed: 22058066]
26. Effemey M, Lang J, Kowalewski J. Multiple-field carbon-13 and proton relaxation in sucrose in viscous solution. *Magnetic Resonance in Chemistry*. 2000; 38:1012–1018.
27. Baraguey C, Mertens D, Dolle A. Anisotropic reorientation and intermolecular interactions of sucrose molecules in aqueous solution. A temperature and concentration-dependent C-13 NMR relaxation study. *Journal of Physical Chemistry B*. 2002; 106:6331–6337.
28. Xia J, Case DA. Sucrose in aqueous solution revisited, Part 1: Molecular dynamics simulations and direct and indirect dipolar coupling analysis. *Biopolymers*. 2012; 97:276–288. [PubMed: 22189655]
29. Lemaster DM, Kay LE, Brunger AT, Prestegard JH. PROTEIN DYNAMICS AND DISTANCE DETERMINATION BY NOE MEASUREMENTS. *Febs Letters*. 1988; 236:71–76. [PubMed: 2841170]
30. James TL, Bianucci AM, Borgias B, Pattabiraman N, Suzuki E, Zhou N, Zon G. MOLECULAR-STRUCTURE DETERMINATION VIA COMPLETE RELAXATION MATRIX ANALYSIS OF 2D NOE SPECTRA AND MOLECULAR MECHANICS - DNA FRAGMENTS. *Biochemistry*. 1987; 26:4157–4158.
31. Suzuki E, Pattabiraman N, Zon G, James TL. SOLUTION STRUCTURE OF D(A-T)5 2 VIA COMPLETE RELAXATION MATRIX ANALYSIS OF TWO-DIMENSIONAL NUCLEAR OVERHAUSER EFFECT SPECTRA AND MOLECULAR MECHANICS CALCULATIONS - EVIDENCE FOR A HYDRATION TUNNEL. *Biochemistry*. 1986; 25:6854–6865. [PubMed: 3801398]
32. Case, DA.; Darden, TA.; Cheatham, ITE.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Walker, RC.; Zhang, W.; Merz, KM.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, AW.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wolf, RM.; Liu, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M-J.; Cui, G.; Roe, DR.; Mathews, DH.; Seetin, MG.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, PA. AMBER. Vol. 12. University of California; San Francisco: 2012.
33. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: A generalizable Biomolecular force field. *Carbohydrates*. *Journal of Computational Chemistry*. 2008; 29:622–655. [PubMed: 17849372]
34. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. *Journal of Chemical Physics*. 1983; 79:926–935.
35. Mahoney MW, Jorgensen WL. Diffusion constant of the TIP5P model of liquid water. *Journal of Chemical Physics*. 2001; 114:363–366.

36. DeMarco ML, Woods RJ. Structural glycobiology: A game of snakes and ladders. *Glycobiology*. 2008; 18:426–440. [PubMed: 18390826]
37. Brown GM, Levy HA. FURTHER REFINEMENT OF STRUCTURE OF SUCROSE BASED ON NEUTRON-DIFFRACTION DATA. *Acta Crystallographica Section B-Structural Science, B*. 1973; 29:790–797.
38. McCain DC, Markley JL. ROTATIONAL SPECTRAL DENSITY-FUNCTIONS FOR AQUEOUS SUCROSE - EXPERIMENTAL-DETERMINATION USING C-13 NMR. *Journal of the American Chemical Society*. 1986; 108:4259–4264.
39. Ortega A, de la Torre JG. Efficient, accurate calculation of rotational diffusion and NMR relaxation of globular proteins from atomic-level structures and approximate hydrodynamic calculations. *Journal of the American Chemical Society*. 2005; 127:12764–12765. [PubMed: 16159246]
40. Battistel MD, Pendrill R, Widmalm G, Freedberg DI. Direct Evidence for Hydrogen Bonding in Glycans: A Combined NMR and Molecular Dynamics Study. *Journal of Physical Chemistry B*. 2013; 117:4860–4869.
41. Teng CL, Hong H, Kiihne S, Bryant RG. Molecular oxygen spin-lattice relaxation in solutions measured by proton magnetic relaxation dispersion. *Journal of Magnetic Resonance*. 2001; 148:31–34. [PubMed: 11133273]
42. Gu Y, Li DW, Brueschweiler R. NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment. *Journal of Chemical Theory and Computation*. 2014; 10:2599–2607. [PubMed: 26580780]

Highlights

- A software package is presented which allows calculation of NOEs from long MD trajectories.
- The package avoids commonly used assumptions about the distance dependence of NOEs.
- Applications to sucrose demonstrate the inseparability of rotational and internal timescales.

Figure 1. Flow chart illustrating the steps leading from selection and validation of an MD trajectory to the calculation of NOE build-up curves.

Figure 2.

Inter-proton distance between fructose H12 (fH12) and glucose H1 (gH1). Distances were extracted every tenth frame of a 100 ns production run for sucrose in a TIP3P water box at 300K. The structures represent major conformers with representative distances.

Figure 3.

Average of the projection of the gH1-fH12 vector on the x-axis of the water box over the trajectory of Figure 1. Convergence to a value near zero supports isotropic tumbling and adequate length of the simulation.

Figure 4. Correlation function calculated from the trajectory of Figure 1. Decay indicates an effective correlation time of about 0.1ns and adequate convergence to zero in 2 ns.

Figure 5. NOE build-up and decay curves for the fH11 and fH12 resonances after saturation of the gH1 resonance. Large diamonds are experimental data collected at 600 MHz and 298 K.

Table 1

Examples of elements from the complete relaxation matrix for sucrose. Entries are in s^{-1} .

Proton pair	ρ (1)	ρ (2)	σ
fH12:gH1	0.99	0.41	0.035
fH11:gH1	1.02	0.41	0.037
fH4:gH1	0.45	0.41	0.003
gH2:gH1	0.44	0.41	0.075

Table 2

Simulated and experimental NOE data along with distance comparisons.

Proton pair	Experiment Initial Slope (s ⁻¹)	Simulation Initial Slope (s ⁻¹)	Experiment rigid r (Å)	Simulation rigid r (Å)	MD <r> (Å)	MD <r> ^{-1/6} (Å)
fH12:gH1	0.041(0.001) **	.035	2.56	2.73	3.03	2.50
fH11:gH1	0.041(0.001) **	.037	2.56	2.70	3.03	2.49
fH4:gH1	0.006(0.0005)	.003	3.52	4.10	4.82	4.17
gH2:gH1	0.061(0.002)	.075	2.4 *	2.4 *	2.39	2.37

* Reference distance.

** These are degenerate resonances and individual slopes cannot be determined; the measured slope has been divided equally between the two pairs. Estimated errors for experimental data are given in parentheses.