WILEY | Hindawi

Research Article

Simulated Annealing Method for the Automated Simulation of DNA Dynamics in the HhaI Protein Binding Site

Kari Pederson (ID),[1] **Michael Groves** (ID),[2] **and Gary P. Drobny** (ID)[3]

[1]*Department of Chemistry & Biochemistry, California State University, Dominguez Hills, Carson, CA 90747, USA*
[2]*Grant NMR Center, University of Utah, Salt Lake City, UT 84112, USA*
[3]*Department of Chemistry, University of Washington, Seattle, WA 98195-1700, USA*

Correspondence should be addressed to Kari Pederson; kpederson@csudh.edu

Solid-state deuterium NMR is well suited to the study of the conformational dynamics of DNA. Deuterium quadrupole echo spectra for a particular motional model can be calculated and matched to the experimental spectrum to extract information on the DNA dynamics; however, doing so can be very time-intensive. The two-axis motion used to model the dynamics of either $2''$ or $5'/5''$ furanose ring deuteron is particularly complex with up to ten independent variables that can be optimized. Here, we present a program which automates both the input script generation and searches the parameter space for the best fit using a simulated annealing algorithm. The parameter, χ^2_{red}, provides a relative measure of goodness of fit. This method reduces the overall time to determine the best fit of a line shape to a few days, in most cases, when running on a low-power desktop PC. The automated fitting program presented here can be easily modified to generate input scripts for new models, incorporate a weighting factor to the χ^2_{red} calculation to emphasize key line shape features, or fit nonsymmetrized data. This adaptable program will make simulation of solid-state deuterium spectra accessible to a broader audience.

1. Introduction

Deuterium solid-state NMR has been widely used in the study of polymer dynamics for many years [1–12] and more recently been used to study dynamics in silica-supported species [13]. It has also been employed to quantitatively evaluate the internal dynamics of the HIV-1 TAR RNA [14, 15] and various DNAs, [16–18] including the HhaI methylase target DNA sequence [19, 20]. Deuterium is particularly well suited to probe nucleic acid dynamics because the ^{2}H label can be incorporated site-specifically and does not perturb the system. Most importantly, it directly monitors the ns-μs regime that is inaccessible to solution relaxation measurements. Both line shapes and relaxation times provide qualitative information about mobility of the ^{2}H label. A quantitative description of the rates and amplitudes of the motion is obtained by comparison of experimental data and model-based simulations of ^{2}H line shapes and relaxation times. In particular, deuterium quadrupole echo spectra are useful for determining motions on multiple timescales or axes. Solid-state ^{2}H NMR is especially effective for observing the large amplitude dynamics within different subunits of individual nucleotides. It is therefore possible to obtain a picture of the relative flexibility at different positions and attempt to quantify the motions therein.

Static ^{2}H NMR has been widely used to study the dynamic properties of solids. In particular, line shape analysis of ^{2}H NMR powder patterns has been widely used and is based on the fact that when a ^{2}H nucleus undergoes reorientational motion that is "intermediate" on the ^{2}H NMR time-scale (i.e. ms-ns), the ^{2}H NMR line shape is altered in a well-defined manner that can be interpreted in terms of the mechanism and rate of the dynamic process [16]. On this time-scale, the ^{2}H NMR line shape is strongly dependent upon the exact rate and geometry of the motion.

Although spectra can be interpreted for qualitative kinetic information, quantitative values can only be obtained

by the matching of calculated spectra to the experimental data. In order to calculate spectra, a detailed motional model (including quadrupole and acquisition parameters and jump rates, amplitudes, and orientations) must be provided. It is generally straightforward to simulate the ^{2}H NMR spectrum theoretically for any proposed dynamic model for a ^{2}H atom.

Typically, the simulated spectrum that best fits each experimental spectrum is found through trial-and-error variation of the model parameters and visual inspection of overlaid simulated and experimental spectra to assess the quality of fit. However, this method is inherently subjective and the large number of parameters which need to be optimized simultaneously in fitting the simulated spectra to the experimental spectra makes this approach highly dependent on the skill of the researcher at recognizing patterns in the effects of various parameters on the ^{2}H line shape. Although it is possible for a researcher to identify trends in line shapes with parameters, the initial time investment of the researcher is on the order of months, and the subsequent fitting of a set of nine line shapes (as illustrated in this paper) would take at least a year. A more objective, efficient, automated, and robust approach for fitting simulated ^{2}H NMR spectra to experimental spectra would clearly be valuable.

Frequently, macromolecules undergo reorientational motions about two or more axes, each characterized by a different trajectory and rate constant. Often small-amplitude librational motions that are rapid with respect to the ^{2}H NMR time-scale must also be included in the model [21–23]. In such cases, line shape analysis is complex, requiring optimization of several independent variables to find the best fit, increasing the time to find the best fit, and reducing the likelihood of successfully finding the best fit to the experimental spectra by the traditional trial-and-error method.

In recent years, several different approaches have been published for removing this subjective aspect from the fitting of solid-state ^{2}H line shapes and relaxation times, both for static and MAS data of peptides, mostly utilizing some form of R^2 or χ^2 analysis [6, 8, 10, 11, 13]. Each of these methods has its own strengths, some combining ^{2}H NMR and other spectroscopic techniques, [5, 24, 25] some incorporating relaxation times into the determination of the best fit, [6, 8, 10, 13] some using static, MAS, and relaxation times together, [9, 12] and some developing automated fitting methods to remove the subjectivity completely [11, 13]. In all cases, these methods are applied to peptides and in some way employ a library of rotamers and/or library of calculated line shapes and relaxation times.

Aliev and Harris have previously applied a simulated annealing and downhill simplex method to the fitting of ^{2}H line shapes of small molecular systems and complexes including thiourea, a thiourea inclusion compound, and pyrazine-d4-α-zirconium intercalation material, with 3 independently varied parameters per reorientational motion, with each of the two possible reorientational motions described by either a 2-site or 4-site jump [26]. Here, we present a modification of this method and apply it instead to DNA motions by utilizing the well-established conformational changes of furanose rings. To illustrate the application of the technique, we consider the dynamic properties of the furanose rings and phosphodiester backbone of the DNA target of the HhaI methyltransferase. Here, we demonstrate how this method can be effectively utilized for the optimization of previously determined best fit simulations (backbone) and also for the determination of best fit simulations starting with at least one parameter changed significantly from previous simulations (furanose ring).

2. Methods

2.1. Experimental. Deuterium labels were incorporated at the 2″ position (furanose ring) of the A4, G5, G7, and A10 and nonstereospecifically into the 5′ methylene group (phosphodiester backbone) of the G5, C6, G7, C8, and 5-methyl-C6 nucleotides in $[d(G_1A_2T_3A_4\underline{G_5}\underline{C_6}\underline{G_7}\underline{C_8}T_9A_{10}T_{11}C_{12})]_2$ (Figure 1) as described previously [19, 20]. Oligonucleotides samples were prepared for NMR analysis and solid-state NMR experiments were carried out at room temperature (298 K) as described previously [19, 20].

2.2. Line Shape Analysis. Simulated line shapes were calculated from parameters describing the global and local motions (Figure 2) using the fitting program described in Section 2.3. To simulate the deuterium line shape, a form of the potential energy surface/landscape, $U(\phi_i)$, (through which the C-^{2}H bond is traversing) must be chosen. The model used to simulate the furanose ring and phosphodiester backbone motions of this DNA sequence has been described in detail previously [19, 20] and is only reviewed briefly here.

As described previously, the local motion is described by a 2-well Brownian diffusion model using a reorientational trajectory originating from the sugar puckering motion of the ring. The reorientational trajectory is approximated using a 10-site jump matrix. The barrier heights of each well of the two-well potential ($U_{0,1}$ and $U_{0,2}$), the amplitude of the motion, and the rate of diffusion are varied during the simulation algorithm. The global motion is described by a 6-site jump diffusion on a cone model, where the orientation of the cone is varied and the rate is set to 10^4 Hz [19, 20]. The central Lorentzian line is a result of residual HDO in the sample and the width of that peak is varied separately during the simulation algorithm. The quadrupolar coupling constant (QCC) was set to 174 kHz for all simulations.

2.3. Fitting Program. The fitting program described here (Figure 3) takes advantage of the existing ^{2}H line shape simulation engine, MXET1, [27] to calculate theoretical ^{2}H line shapes. Initial "good" fits have been found previously using traditional visual inspection of a library of simulated line shapes [19, 20]. To refine these fits and find the best fit simulated line shape, an algorithm for searching the parameter space and a cost function to quantify the goodness of fit must be defined. An "input script generator" creates an MXET1 input script for each new step in the algorithm using maximum step sizes and boundary values for each variable specified by the researcher.
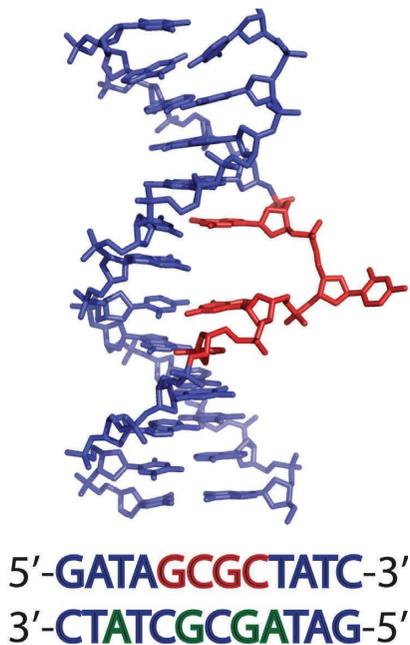
FIGURE 1: (Top) structure of the dodecamer containing the HhaI target GCGC sequence (red) in the protein-DNA complex (protein removed), showing the extrahelical deoxycytidine. (Bottom) sequence of the target DNA; residues highlighted in red are labeled at the $5'/5''$ position. Residues highlighted in green are labeled at the $2''$ position.

2.3.1. Simulated Annealing Algorithm. Simulated annealing mimics the physical annealing of a solid, in which the solid is heated and then slowly cooled to avoid local high-energy configurations and reach the lowest energy state [28, 29]. Typically, in this procedure, a cost function (analogous to energy) is minimized while the value of a control parameter (analogous to temperature) is decreased. However, here, the control parameter is held constant as it has been shown that if the control parameter is chosen carefully, it is not necessary to decrease the value over time [30]. As discussed previously, in order to obtain satisfactory agreement between experimental and simulated ^{2}H NMR spectra, it is necessary to optimize several parameters that define the dynamic model. Following the method of Aliev and Harris, [26] each value of I_i^{calc} can be thought of as a function of the parameters that are refined during the fitting process. The automated fitting procedure employed here adjusts the parameters in order to minimize the cost function (S), indicating to the best fit between experimental and simulated spectra [31]. The condition for best fit is defined here as 1000 simulations without an improved S value.

The simulated annealing algorithm explores parameter space following the method published by Metropolis, et al. [32] summarized as follows. First, an initial set of parameters are chosen. Second, a new parameter set is derived from the previous parameter set via two steps.

(1) From a chosen starting configuration (set of parameters), a random step is made; i.e., each parameter within the set is randomly "displaced" to generate a new "trial" value, giving rise to a new trial

parameter set. A maximum absolute displacement value is specified for each parameter value.

(2) The value of S is then calculated; if the step lowers the value of the cost function, it is always accepted; otherwise, the step is accepted when a randomly chosen probability ($0 < P < 1$) satisfies the following condition:

$$P \le \exp\left(-\frac{\Delta S}{c}\right). \quad (1)$$

Here, c is the control parameter and ΔS is the difference between the current and previous values of the cost function. As mentioned previously, in this study, the control parameter was held constant. The nonzero probability of accepting an uphill step ensures that it is possible to escape from local minima, increasing the likelihood of finding the global minimum of the cost function.

Steps 1 and 2 are repeated to generate a Markov chain of parameter sets. By requiring 1000 steps without improvement, this technique leads, in principle, to the region of parameter space containing the lowest minimum in the merit function S. In our calculations, c was chosen to be 10.

The approach of Aliev and Harris [26] was to calculate spectra during the fitting process. For our DNA models where we had previously determined good fits through manual trial and error and visual comparison, we have also chosen to calculate the spectra during the fitting process. To reduce the number of external programs called during the simulation algorithm, a library of 10 site angular trajectories for amplitudes ranging from 0.001 Å to 0.800 Å was generated in advance for the program to use when creating the input scripts.

2.3.2. The Cost Function. Aliev and Harris [26] developed a method of automated fitting of quadrupole echo spectra for a given motional model using the method of simulated annealing, 28 with a cost function R,

$$R = \frac{1}{N} \sum_{i=1}^{N} \left(I_i^{\text{exp}} - I_i^{\text{calc}} \right)^2, \quad (2)$$

where I^{exp} is the normalized experimental spectrum to be fit, I^{calc} is the calculated spectrum, and N is the number of points in the spectral pattern. Here, we have used a cost function S and is given as

$$S = \sum_{i=1}^{N} \left(I_i^{\text{exp}} - I_i^{\text{calc}} \right)^2, \quad (3)$$

which is not reduced by the number of points in the spectral pattern and therefore is more sensitive to slight improvements in the quality of the fit. The experimental spectrum is normalized to a maximum intensity of 100 to facilitate direct comparison with simulated spectra. In addition, the experimental spectrum is centered and the simulated spectra contain the same number of data points as the experimental

(a)   (b)
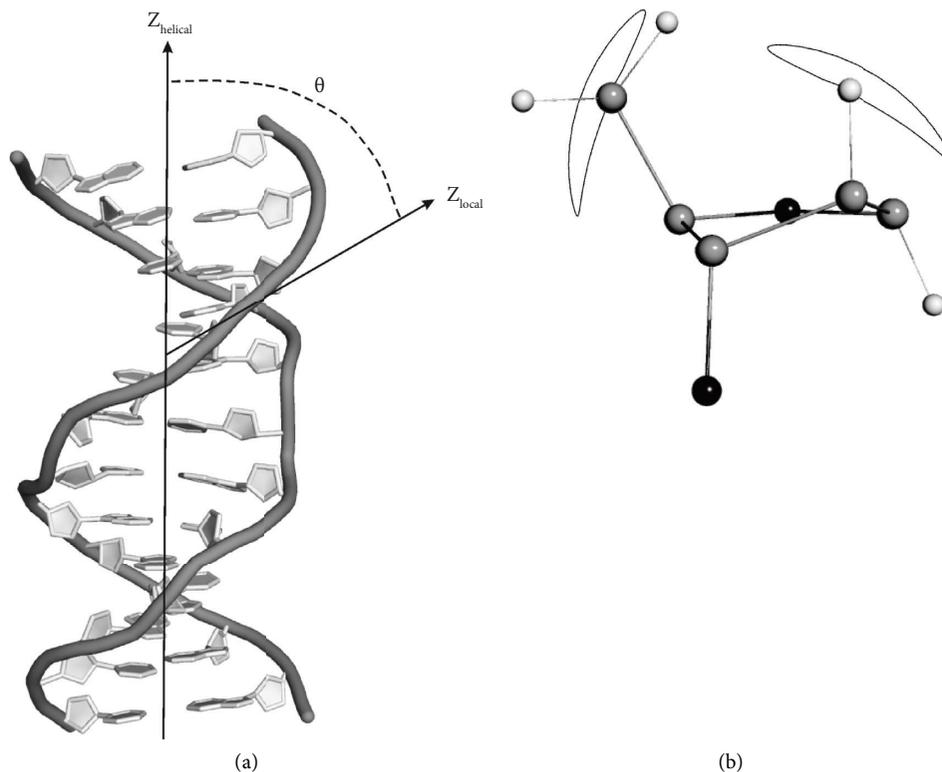
FIGURE 2: Line shape simulations include two independent motions. (a) A slow, uniform rotation of the DNA molecule occurs about the helical axis, labeled Z_{helical}. Local motion of the backbone is referenced to a local coordinated system, where the z axis is indicated by the vector Z_{local} defined by the angle Θ. (b) The $^2\text{H}_{5'/5''}$ and $^2\text{H}_{2''}$ deuterons move through an elliptical trajectory as the furanose ring puckers between the $C_{2'}$-endo and $C_{3'}$-endo conformations.

data such that the points match for evaluation of the cost function.

For comparison of goodness of fit across many different spectra, a reduced χ^2 must be calculated as follows:

$$\chi^2_{\text{red}} = \frac{1}{N}\frac{S}{\sigma^2}, \tag{4}$$

where the noise standard deviation, σ, is calculated for each spectrum as the standard deviation of the amplitude of all baseline points and N is the number of points (or degrees of freedom), as defined earlier. The reduced χ^2 removes the variability in S arising from any difference in the number of points, noise, and other error levels in each spectrum and can be used as a universal criterion for evaluation of goodness of fit.

3. Results

We have previously reported experimental line shapes, relaxation times, and simulations for the phosphodiester backbone and furanose ring of the 5'-GCGC-3' site. [19, 20] To probe local mobility in this site, deuterium line shapes were obtained for the 5'/5'' deuterons (backbone) of G5, C6, G7, C8, and 5-methyl-C6,[20] and 2'' deuteron (furanose ring) of A4, G5, G7, and A10 [19]. Experimental line shapes and line shape simulations for the backbone and furanose ring sites are shown in Figures 4 and 5, respectively. These line shape results indicated that the backbone of C6 is

significantly more mobile than the other nucleotide backbones on the ns-ms timescale, followed by G5 and G7. Previous manually determined line shape simulations quantified and supported these conclusions, demonstrating an increased flexibility and distinctive orientation of the C6 backbone. The fitting procedure described previously was applied to analyze the ^2H NMR spectra of these backbone and furanose ring labeled DNAs.

3.1. Analysis of Backbone Line Shapes. As mentioned in Section 2.2 and described previously, it was assumed that as a result of coupling to the furanose motion, the reorientational motion of the 5' methylene C-^2H bonds is not simple rotations about the $C_{4'}$-$C_{5'}$ bond axes. Simulated line shapes in Figure 4 were obtained by setting values for the jump rate k, the displacement amplitude, and the potential $U(\phi)$. For [5'/5''-^2H]-G5, [5'/5''-^2H]-C6, [5'/5''-^2H]-G7, and [5'/5''-^2H]-C8, best fits are obtained for $U(\phi)$ modeled as a periodic potential of the form

$$U(\phi) = \frac{U_0}{2}\left(1 - \cos\left(\phi - \phi_0\right)\right), \tag{5}$$

where U_0 is the barrier between the minimum energy conformers (in this case the B_{I} and B_{II} backbone conformations) and can be varied for each barrier individually. According to (5), the conformational energy displays two wells within the pseudorotation cycle, which occur at $\phi -$
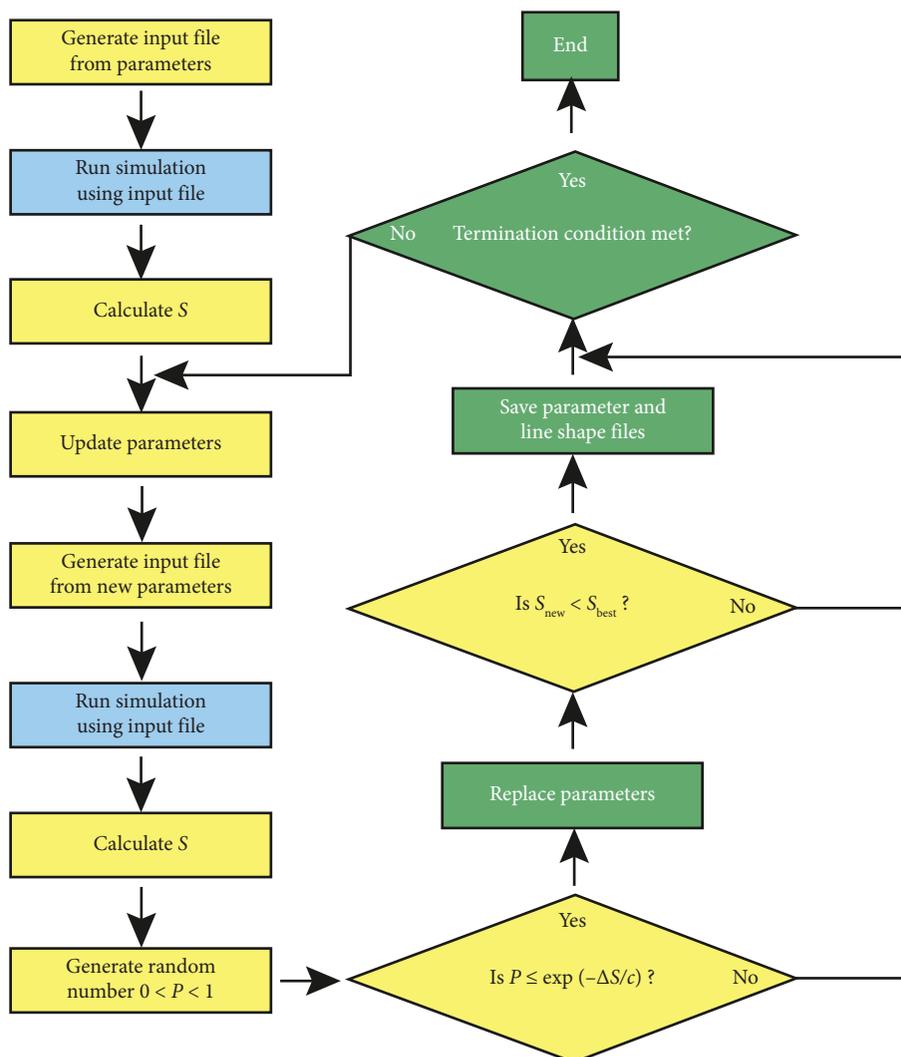
FIGURE 3: Simulated annealing algorithm flow chart. Elements of preexisting Fortran code are shown in blue, while the new C++ code is shown in yellow, and the Microsoft.bat file sections are shown in green.

$\phi_0 = 0°$ and $\phi - \phi_0 = 180°$. The parameters describing the best-fit line shapes and reduced χ^2 (χ^2_{red}) values are shown in Table 1.

We have previously reported manual best fit simulations [20] and here report the optimized best fits determined through use of our automated simulated annealing program. In all but one case (G7), the automated fitting resulted in both a better χ^2_{red} value and a visually assessed equivalent or better fit. Although the automated fit of the G7 line shape had a lower χ^2_{red}, upon visual inspection, the manual simulation fits the major features at the center of the line shape better. While the specific values of the parameters differ between the manual best fits and the overall best fits, the overall trends remain the same. In general, the angular displacement, q, for the best automated fits were consistent with those from the best manual fits. The barrier heights, U_0, for G5, C6, and G7 decreased during the automated fitting, while those for C8 stayed the same and equalized for the mC6. The z_{local} axis angle relative to $z_{helical}$ increased slightly

for the automated fits of G5 and C8, increased dramatically for G7, and decreased for mC6.

3.2. Analysis of Furanose Ring Line Shapes.
Simulated line shapes in Figure 5 were obtained by setting values for the jump rate k, the puckering amplitude q, and the potential $U(\phi)$. For [2″-^{2}H]-A4, [2″-^{2}H]-G5, [2″-^{2}H]-G7, and [2″-^{2}H]-A10, best fits are obtained for $U(\phi)$ modeled as a periodic potential of the form given in (5), where U_0 is the barrier between the minimum energy conformers (in this case, the $C_{2'}$-endo and $C_{3'}$-endo furanose ring conformations) and can be varied for each barrier individually. As before, the conformational energy displays two wells within the pseudorotation cycle, which occur at $\phi - \phi_0 = 0°$ and $\phi - \phi_0 = 180°$. Table 2 summarizes these parameters for the line shapes that best fit the experimental line shapes for [2″-^{2}H]-A4, [2″-^{2}H]-G5, [2″-^{2}H]-G7, and [2″-^{2}H]-A10 and gives the reduced χ^2 (χ^2_{red}) values for each fit.
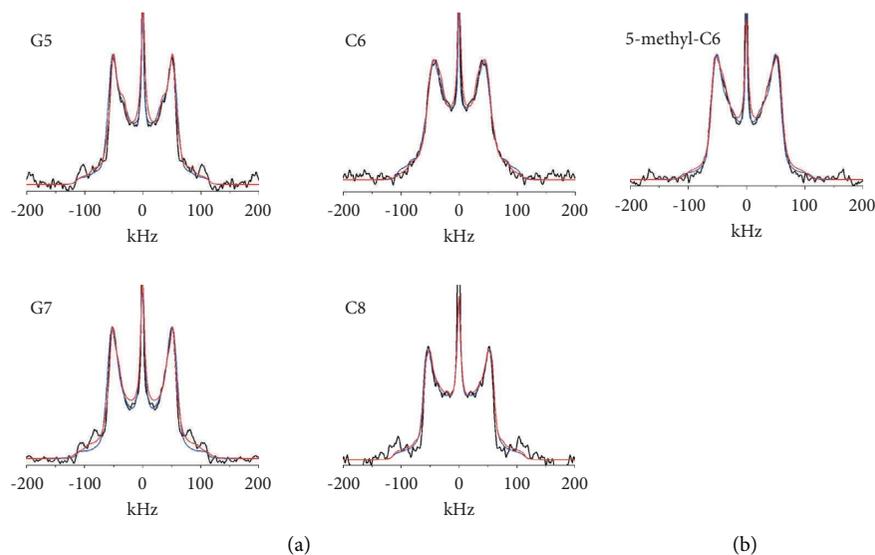
FIGURE 4: The experimental (black) deuterium line shapes with the manual simulation (blue) and automated simulation (red) overlaid for (a) each of the backbone labeled sites in the nonmethylated DNA dodecamer and (b) the $C_{5'/5''}$-^{2}H in the backbone of methylated C6. Simulation parameters are described in Table 1.
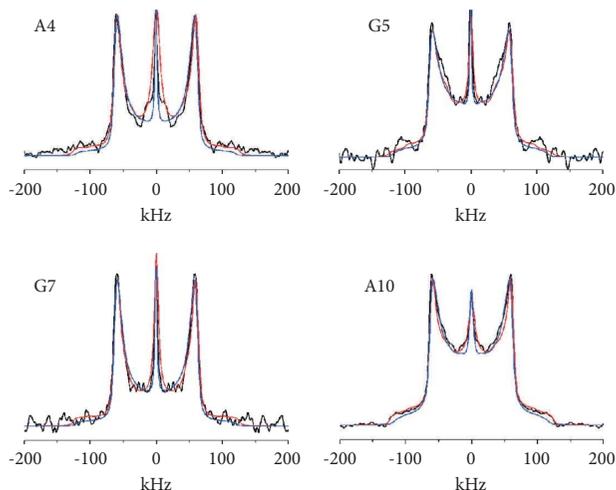
FIGURE 5: The four deuterium line shapes (black) for each of the purine furanose ring labeled sites in the nonmethylated DNA dodecamer with the manual simulation (blue) and automated simulation (red) of each overlaid. Simulation parameters are described in Table 2.

We have previously reported manual best fit simulations [19] and here report the optimized best fits determined through use of our automated simulated annealing program. The manual fits used an angle of 20° for the orientation of the local axis with respect to the helical axis. However, this orientation was based on the average orientation of the 2″ C-D bond, whereas the pseudorotation trajectory uses an axis approximately parallel to the $C_{2'}$-$C_{3'}$ bond, determined to be on average 70°. An orientation of 70° was used for the subsequent automated line shape simulations. The puckering amplitudes, q, for the best manual fits for these four nucleotide furanose rings are approximately 0.25 Å, while the barrier heights, U_0, range from 4.5 k_BT to 6 k_BT.

In contrast, the best fits of the automated simulations for these four nucleotide furanose rings give puckering amplitudes, q, in the range of 0.16 Å to 0.23 Å, and barrier heights, U_0, ranging from 2.3 k_BT to 8.9 k_BT.

3.3. Impact of C6 Methylation on Dynamics.

Methylation of the target cytidine base in the 5′-GCGC-3′ sequence of perturbs the dynamics of the phosphodiester backbone at C6 [20]. As has been discussed previously, there is a significant change in the line shape of [5′/5″-^{2}H]-C6 upon methylation (Figure 4), as the horns expand and "unbend" approaching the classical Pake doublet line shape indicating that this site becomes more rigid upon methylation [20].

Both manual and automated fits indicate that the C6 backbone line shape undergoes significant changes upon methylation. In both cases, upon methylation, the barrier heights increased and the jump rate decreased, although the decrease in rate for the automated fits was a factor of 2 smaller than that of the manual fits. Interestingly, the automated fits showed that the amplitude of motion increased slightly upon methylation, rather than decreasing to become more similar to the other nucleotides. Finally, the orientation of z_{local} decreased to 20° in the automated fits, rather than 40° from the manual fits, indicating an even more pronounced conformational change in the backbone of the C6 nucleotide upon methylation than was indicated with the manual simulations.

This indicates that the C6 5′/5″ deuterons are almost equally likely to be in the two most favored conformations, that they are moving through a larger volume in space than the neighboring sites even after methylation but at a slower rate than all but the G7 5′/5″ deuterons, and that the backbone orientation changes significantly upon methylation. Each of these parameters support the conclusions drawn from the qualitative analysis, namely, that there is

TABLE 1: The potential, given in equation (5), was used for these simulations of the phosphodiester backbone. The values for U_0 along with other parameters used to fit each of the 5 simulated line shapes in Figure 4 are given.

Label site	Fitting method	$U_{0,1}$ (k$_B$T)	$U_{0,2}$ (k$_B$T)	Amplitude (Å)	Jump rate (Hz)	θ	χ^2_{red}
G5	Manual	2.0	4.5	0.30	1×10^8	50°	1.73
	Automated	2.0	3.7	0.31	2.1×10^8	56°	1.13
C6	Manual	1.5	1.5	0.37	1×10^9	60°	1.66
	Automated	0.75	1.0	0.36	1.7×10^8	61°	1.31
G7	Manual	4.0	8.0	0.32	1×10^8	40°	8.44
	Automated	1.9	1.5	0.27	8.8×10^7	71°	3.91
C8	Manual	1.0	3.0	0.32	1×10^8	30°	2.67
	Automated	0.9	3.1	0.33	1.4×10^8	35°	2.46
5mC6	Manual	2.5	4.5	0.32	1×10^8	40°	2.74
	Automated	3.6	3.7	0.39	1.2×10^8	20°	1.82

TABLE 2: The potential, given in equation (5), was used for these simulations of the purine furanose rings. The values for U_0 along with other parameters used to fit each of the 4 simulated line shapes in Figure 5 are given.

Label site	Fitting method	$U_{0,1}$ (k$_B$T)	$U_{0,2}$ (k$_B$T)	Amplitude (Å)	Jump rate (Hz)	θ	χ^2_{red}
A4	Manual	6.0	6.0	0.25	2×10^8	20°	7.85
	Automated	8.8	6.7	0.16	1.8×10^8	61°	3.53
G5	Manual	4.5	4.5	0.27	3×10^8	20°	1.32
	Automated	1.5	2.3	0.17	8.6×10^7	60°	1.02
G7	Manual	6.0	6.0	0.25	2×10^8	20°	2.14
	Automated	5.1	15.3	0.18	7.2×10^7	75°	1.17
A10	Manual	4.5	4.5	0.27	1×10^9	20°	14.0
	Automated	0.19	0.13	0.19	9.7×10^7	60°	3.30

a dramatic increase in rigidity upon methylation of C6, and they demonstrate that upon methylation, the dynamics of the target nucleotide are diminished and become similar to the nontarget sites in the recognition sequence.

4. Discussion

It is not possible to fully automate the determination of the simulated spectrum which best matches an experimental spectrum because it is necessary to subjectively evaluate the spectrum in consideration of the physical system in order to determine a motional model and reasonable limits on the parameters which define that model. It is necessary to consider all physically plausible models and ranges of parameters in the search for a best fit. We have demonstrated the success of the program presented here as a tool to facilitate this search when spectra represent a specified motional model, which consists of torsional motions around multiple axes in the local structure of a DNA methyltransferase recognition site.

Our examples of fitting are restricted to only spectra representing intermediate motions, which are sensitive to rates of motion as well as amplitudes of motion; the appropriate models have rates of 5×10^7 Hz to 1×10^9 Hz. The fitting program is designed to extract rates, amplitudes, and relative orientations of motion from spectral fits. Here, we have shown that the simulated annealing algorithm can be successfully applied both to the optimization of previously determined line shape simulations as well as to the determination of new line shape

simulations for a system with one parameter which required significant adjustment.

Eastman and Nanny utilized a simulated annealing algorithm to search through a set of simulation models and over a defined parameter space to find good fits to the ^{2}H ssNMR spectra of several model compounds[33] and, as such, utilized a more stringent goodness-of-fit function, Q. Here, for a more complex system, we have shown that χ^2_{red} can, in principle, be an effective goodness-of-fit parameter, inasmuch as it is helpful in identifying the best fit for a given model and parameter space. Because of the complex motions of the C-^{2}H bond within the DNA molecule, it may be unrealistic to expect to be able to consistently identify a fit that would be classified as good using either the function Q (a value close to 0.5) or χ^2_{red} (a value between 1 and 1.5).

In addition, in some cases, such as that of the [5′/5″-^{2}H]-G7, the best χ^2_{red} value may correspond to a line shape simulation which does not subjectively fit the key features of the experimental line shape as well as a simulation with a higher χ^2_{red} value, indicating that not all need for subjective evaluation has been removed. The failure to find fits to experimental spectra both with low χ^2_{red} values and visually determined to be acceptable may also be due to experimental artifacts which may have a larger effect on the edges of the line shape and are not taken into account in the quadrupole echo spectral calculation. Furthermore, to facilitate fitting of our experimental line shapes, they have been artificially symmetrized, which effectively reduces the magnitude of experimental artifacts but also makes it more difficult to distinguish artifacts from the line shape. Some researchers

have previously suggested that more accurate simulations may be achieved by fitting only one half of the line shape, rather than artificially symmetrizing the line shape which can inadvertently intensify the effects of artefacts. In this case, the half of the line shape to be evaluated would be chosen to have the fewest artifacts subjectively. The program presented here could be easily modified to this approach.

In the particular case of the line shape of the $2''$-A10 deuteron, where the spectrum itself appears free of obvious artifacts and has good signal to noise, it is noteworthy that the χ^2_{red} value does improve for the automated fit but is still quite high. For line shapes such as this, the high χ^2_{red} value could be evidence of a negative effect from symmetrization or that an additional reorientational motion is missing from the model. When fitting only the line shape, we are reluctant to add an additional axis of motion to the model, which may overdefine the system. However, future modeling that includes the T_{1z} and T_{1Q} relaxation simulations could provide more support for adding a third axis of motion.

While the exact values of the parameters for the best fit simulations of the backbone line shapes differ between our previously reported results [20] and those found using the automated method reported here, the trends remain the same. For example, the $5'$ methylene group of the C6 is the most flexible of the sites studied, with the largest angle between the global and local axes, which decreases significantly (to become more similar to the other deoxycytidine nucleotide) upon methylation. In addition, the purine furanose rings were simulated with a two-well potential with small-amplitude puckering motions. The depths of the wells changed in each case, but A4 and G7 were still fit best using deeper wells than G5 and A10.

5. Conclusion

Here, we have demonstrated the use of our simulated annealing algorithm for simulating line shapes for the methylene groups within the phosphodiester backbone and the deuterons at the $2''$ position of the furanose ring within and adjacent to the DNA recognition site for the HhaI methyltransferase. A visual inspection of the nine simulated line shapes indicated that the cost function, S, provides a good measure of relative goodness of fit for a given line shape and that the reduced χ^2 (χ^2_{red}) can be used to compare fits for line shapes from several different samples.

In some cases, experimental artifacts can affect the goodness-of-fit parameters and lead to "best fits" which upon visual inspection are inferior to other simulations. Future program development might include a weighting parameter to allow the program to account for apparent artifacts in the experimental data by giving priority to the fitting of specific regions of the line shape or fitting only one half of the original unsymmetrized line shape.

Data Availability

Previously reported solid-state NMR data were used to support this study and are available at https://doi.org/10.1021/ja075775n and https://doi.org/10.1021/ja801243d.

These prior studies (and datasets) are cited at relevant places within the text as references [25, 26]. The source code for the program developed in this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Hentschel, J. Schlitter, H. Sillescu, and H. W. Spiess, "Orientational distributions in partially ordered solids as determined from NMR and ESR line shapes," *The Journal of Chemical Physics*, vol. 68, no. 1, pp. 56–66, 1978.

[2] D. Hentschel, H. Sillescu, and H. W. Spiess, "Molecular motion in solid polyethylene as studied by 2D wide line NMR spectroscopy," *Makromolekulare Chemie*, vol. 180, no. 1, pp. 241–249, 1979.

[3] H. W. Spiess, "Deuteron spin alignment: a probe for studying ultraslow motions in solids and solid polymers," *The Journal of Chemical Physics*, vol. 72, no. 12, pp. 6755–6762, 1980.

[4] H. Menge, S. Hediger, L. Emsley, and D. Sandström, "Segmental mobility in poly(isoprene) rubber studied by deuterium-carbon NMR correlation spectroscopy," *Polymer Bulletin*, vol. 46, no. 2-3, pp. 183–190, 2001.

[5] B. Reif, Y. Xue, V. Agarwal et al., "Protein side-chain dynamics observed by solution- and solid-state NMR: comparative analysis of methyl 2H relaxation data," *Journal of the American Chemical Society*, vol. 128, no. 38, pp. 12354-12355, 2006.

[6] L. Vugmeyster, D. Ostrovsky, M. Moses et al., "Comparative dynamics of leucine methyl groups in FMOC-leucine and in a protein hydrophobic core probed by solid-state deuteron nuclear magnetic resonance over 7–324 K temperature range," *Journal of Physical Chemistry B*, vol. 114, no. 48, pp. 15799–15807, 2010.

[7] J. R. Long, F. D. Mills, O. K. Ganesh, V. C. Antharam, and R. S. Farver, "Partitioning, dynamics, and orientation of lung surfactant peptide KL4 in phospholipid bilayers," *Biochimica et Biophysica Acta (BBA) Biomembranes*, vol. 1798, no. 2, pp. 216–222, 2010.

[8] L. Vugmeyster, D. Ostrovsky, K. Penland, G. L. Hoatson, and R. L. Vold, "Glassy dynamics of protein methyl groups revealed by deuteron NMR," *Journal of Physical Chemistry B*, vol. 117, no. 4, pp. 1051–1061, 2013.

[9] K. Li, P. S. Emani, J. Ash, M. Groves, and G. P. Drobny, "A study of phenylalanine side-chain dynamics in surface-adsorbed peptides using solid-state deuterium NMR and rotamer library statistics," *Journal of the American Chemical Society*, vol. 136, no. 32, pp. 11402–11411, 2014.

[10] L. Vugmeyster, D. Ostrovsky, T. Villafranca et al., "Dynamics of hydrophobic core phenylalanine residues probed by solid-state deuteron NMR," *Journal of Physical Chemistry B*, vol. 119, no. 47, pp. 14892–14904, 2015.

[11] E. Meirovitch, Z. Liang, and J. H. Freed, "Protein dynamics in the solid state from 2H NMR line shape analysis: a consistent

perspective," *Journal of Physical Chemistry B*, vol. 119, no. 7, pp. 2857–2868, 2015.

[12] N. F. Breen, K. Li, G. L. Olsen, and G. P. Drobny, "Deuterium magic angle spinning NMR used to study the dynamics of peptides adsorbed onto polystyrene and functionalized polystyrene surfaces," *Journal of Physical Chemistry B*, vol. 115, no. 30, pp. 9452–9460, 2011.

[13] H. E. Ferreira and G. P. Drobny, "Solid state deuterium NMR study of LKα14 peptide aggregation in biosilica," *Biointerphases*, vol. 12, no. 2, pp. 411–410, Article ID 02D418, 2017.

[14] G. L. Olsen, D. C. Echodu, Z. Shajani, M. F. J. Bardaro, G. Varani, and G. P. Drobny, "Solid-state deuterium NMR studies reveal μs-ns motions in the HIV-1 transactivation response RNA recognition site," *Journal of the American Chemical Society*, vol. 130, no. 10, pp. 2896-2897, 2008.

[15] W. Huang, P. S. Emani, G. Varani, and G. P. Drobny, "Ultraslow domain motions in HIV-1 TAR RNA revealed by solid-state deuterium NMR," *Journal of Physical Chemistry B*, vol. 121, no. 1, pp. 110–117, 2016.

[16] T. M. Alam and G. P. Drobny, "Solid-state NMR studies of DNA structure and dynamics," *Chemistry Review*, vol. 91, no. 7, pp. 1545–1590, 1991.

[17] M. E. Hatcher, D. L. Mattiello, G. A. Meints, J. Orban, and G. P. Drobny, "A solid-state deuterium NMR study of the localized dynamics at the C9pG10 step in the DNA dodecamer [d(CGCGAATTCGCG)]₂," *Journal of the American Chemical Society*, vol. 120, no. 38, pp. 9850–9862, 1998.

[18] K. B. Geahigan, G. A. Meints, M. E. Hatcher, J. Orban, and G. P. Drobny, "The dynamic impact of CpG methylation in DNA," *Biochemistry*, vol. 39, no. 16, pp. 4939–4946, 2000.

[19] G. A. Meints, P. A. Miller, K. Pederson, Z. Shajani, and G. P. Drobny, "Solid-state nuclear magnetic resonance spectroscopy studies of furanose ring dynamics in the DNA HhaI binding site," *Journal of the American Chemical Society*, vol. 130, no. 23, pp. 7305–7314, 2008.

[20] K. Pederson, G. A. Meints, Z. Shajani, P. Miller, and G. P. Drobny, "Backbone dynamics in the DNA HhaI protein binding site," *Journal of the American Chemical Society*, vol. 130, no. 28, pp. 9072–9079, 2008.

[21] R. J. Schadt, E. J. Cain, and A. D. English, "Simulation of one-dimensional deuteron NMR line shapes," *Journal of Physical Chemistry*, vol. 97, no. 32, pp. 8387–8392, 1993.

[22] J. H. Ok, R. R. Vold, R. L. Vold, and M. C. Etter, "Deuterium NMR measurements of rotation and libration of benzene in a solid-state cyclamer," *Journal of Physical Chemistry*, vol. 93, no. 22, pp. 7618–7624, 1989.

[23] S. J. Heyes and C. M. Dobson, "13C cross-polarization/magic angle spinning and 2H NMR studies of the structure and dynamics of the (deoxycholic acid)2–ferrocene inclusion compound," *Magnetic Resonance in Chemistry*, vol. 28, no. 13, pp. S37–S46, 1990.

[24] T. Weidner, N. F. Breen, K. Li, G. P. Drobny, and D. G. Castner, "Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 30, pp. 13288–13293, 2010.

[25] K. P. Howard, W. Liu, E. Crocker et al., "Rotational orientation of monomers within a designed homo-oligomer transmembrane helical bundle," *Protein Science*, vol. 14, no. 4, pp. 1019–1024, 2005.

[26] A. E. Aliev and K. D. M. Harris, "2H NMR lineshape analysis using automated fitting procedures based on local and quasi-global optimization techniques," *Magnetic Resonance in Chemistry*, vol. 36, no. 11, pp. 855–868, 1998.

[27] R. R. Vold and R. L. Vold, "Deuterium relaxation in molecular solids," *Advances in Magnetic and Optical Resonance*, vol. 16, pp. 85–171, 1991.

[28] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[29] P. J. M. V. Laarhoven and E. H. L. Aarts, *Simulated Annealing Theory and Applications*, D Reidel, Dordrecht, Netherlands, 1987.

[30] D. T. Connolly, "An improved annealing scheme for the QAP," *European Journal of Operational Research*, vol. 46, no. 1, pp. 93–100, 1990.

[31] S. Kirkpatrick, "Optimization by simulated annealing: quantitative studies," *Journal of Statistical Physics*, vol. 34, no. 5-6, pp. 975–986, 1984.

[32] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[33] M. A. Eastman and M. A. Nanny, "Fitting of deuterium quadrupole echo spectra with multiple motional models," *Journal of Magnetic Resonance*, vol. 184, no. 2, pp. 302–314, 2007.