Chapter 2

USING AMBER TO SIMULATE DNA AND RNA

Thomas E. Cheatham, III[1] and David A. Case[2]
[1]Departments of Medicinal Chemistry and of Pharmaceutics and Pharmaceutical Chemistry and of Bioengineering, 2000 East, 30 South, Skaggs Hall 201, University of Utah, Salt Lake City, UT 84112, USA, tec3@utah.edu; [2]Department of Molecular Biology, TPC15, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037, USA, case@scripps.edu

Abstract: The use of molecular dynamics and free energy simulation model nucleic acid structure, dynamics, and interactions are discussed from the authors AMBER-centric viewpoint.

Key words: AMBER; molecular dynamics; DNA; RNA; biomolecular simulation; MM-PBSA; force fields; Ewald

1. AMBER APPLIED TO NUCLEIC ACIDS

AMBER, "Assisted Model Building with Energy Refinement", is a suite of programs that has evolved over the past three decades to enable the application of molecular dynamics (and free energy) simulation methods to bio-molecular systems including nucleic acids, proteins, and more recently carbohydrates. Starting with the first simulations of DNA in explicit water[1] and progressing to recent large-scale explicit solvent simulations of components of the ribosome[2, 3] [and also see Chapter 12], AMBER or its associated force fields have arguably been applied in the majority of published simulations involving nucleic acids to date. This relates to the performance of the AMBER-related force fields[4-6] (and their free availability) and the relatively early adoption of fast and efficient particle mesh Ewald methods[7, 8] for the proper treatment of long-range electrostatic interactions. These advances allowed stable simulation of nucleic acids in explicit solvent[9, 10] including an accurate representation of the surroundings

J. Šponer and F. Lankas (eds.), Computational Studies of RNA and DNA, 45–71.

and sequence dependent structure and dynamics[11-18]. More recently — and in addition to the extensive and emerging sets of simulation in explicit solvent—there has been resurgence in the application of generalized Born implicit solvent methods[19-21] in nucleic acid simulation[22]. This has been facilitated by the development of better methods and force fields in both CHARMM[23] and AMBER[24].

In this chapter we highlight the use of molecular dynamics simulation applied to nucleic acids from our limited author- and AMBER-centric viewpoint. For more thorough background related to AMBER development and application, see related review articles[24, 25]. For a complete review of MD methods applied to nucleic acids and their success, a number of comprehensive reviews have been published[26-31]. For a more complete review of nucleic acid force fields and their performance, see our work[32], or work by MacKerell[33], Karplus[34], Langley[35], van Gunsteren[36], or also the ABC consortium[18]. In addition, a set of guides to nucleic acid simulation have been published that may help introduce the field and methods[37]. To aid in learning and using AMBER, a number of tutorials (which include applications to nucleic acids) are distributed with the program and available on the WWW (see http://amber.scripps.edu).

It is useful to note that AMBER has not been developed or designed as a program that can do everything or that aims to include *all* molecular dynamics methods. Instead it is focused into a series of specialized programs that allow set-up and limited model building (LEaP, `antechamber`), accurate and efficient molecular dynamics and free energy simulation (`sander`, `PMEMD`) and analysis (`ptraj`, `carnal`, `mm-pbsa`). A brief introduction to each of these general capabilities (in the context of nucleic acid simulation) is provided below. AMBER is also not a force field, although a number of force fields developed in the developer's and other research labs are distributed with the suite of programs.

1.1 Setting up AMBER MD Simulations

LEaP is a freely available (X windows or text based) program that facilitates setting up the parameter/topology and coordinate input files for subsequent MD studies. Only very limited model building is provided and the normal starting point is the reading of specific force field specifications (for a variety of available force fields) and coordinates (in the form of a PDB file). Since AMBER 7.0, the `antechamber` program (also freely available) has been distributed. This program, in addition to facilitating file conversion from Gaussian and other file formats into those readable by LEaP, includes access to the GAFF (general AMBER force field) force field[38] that is

intended to provide a means to specify decent molecular mechanical parameters for drug-like molecules or new residues. Details into its power and applicability are described elsewhere[39, 40]; the point of discussing the program here is to point out that now it is significantly easier to model DNA modifications or nucleic acid-ligand interactions than previously since the difficult part of obtaining parameters for a new residue is greatly simplified. Both of these programs do not provide significant model building capability and therefore users typically rely on experimental structures as starting points or must generate their own reliable model structures. To generate arbitrary nucleic acid models, NAB (nucleic acid builder) is a general purpose molecular manipulation programming language that has been developed[41, 42]. It is intended for the building of both helical and non-helical models with up to 100s of nucleotides using a combination of rigid body transformations and distance geometry; the examples supplied with the program show how to build a DNA mini-circle such as is discussed later in this chapter. The program can also perform molecular mechanics including generalized Born implicit solvent methods. Other common programs used for building nucleic acid models include 3DNA[43], ERNA-3D[44, 45], MANIP[46], MC-SYM[47], and NAMOT[48].

1.2 AMBER Dynamics

The main molecular dynamics engine is the `sander` program which is focused primarily in two directions: (1) accurate and parallel efficient simulation of explicitly solvated periodically replicated unit cells and (2) fast/efficient simulation of non-periodic systems in implicit solvent. Development is driven largely by research and the needs of the associated developers; yet, despite being primarily a research code, the programs are widely utilized throughout the world. Specialized methods that are highlighted in `sander` include a general facility for including distance, angle, torsion, NOESY volume, chemical shift, and residual dipolar coupling restraints, an evolving efficient QM/MM implementation[49, 50], a specially optimized parallel version of the PME code (PMEMD by Robert Duke), a new long-range electrostatics methods developed by Wu called isotropic periodic (IPS) boundary conditions[51], an efficient energy-mixing based thermodynamic integration facility, and various enhanced sampling methods. The enhanced and/or path sampling methods include:

- Umbrella sampling using the NMR-based restraints in `sander`.
- LES: locally enhanced sampling[52, 53] is a method which lowers barriers to conformational exchange by generating copies of sub-regions of a molecule that are seen by the rest of the molecule as an average.

- Replica-exchange: Two facilities, either internally or externally via the MMTSB toolkit[54], are provided which allow simultaneous running of separate calculations with exchange of information (such as temperature).
- Targeted-MD: Restraints have been added that use the RMSd (best fit root-mean-squared deviation) as a reaction coordinate.

As shown consistently in the literature, an accurate description of the aqueous environment is essential for realistic bio-molecular simulations. However, this easily becomes very expensive computationally. For example, an adequate representation of the solvation of a medium-size protein typically requires thousands of discrete water molecules to be placed around it. While some of the cost is ameliorated by using fast efficient methods for treating the long-ranged electrostatics (such as PME[8, 55] or IPS) and specially optimized code like PMEMD, an alternative replaces the discrete water molecules by "virtual water"– an infinite continuum medium with the (some of the) dielectric and hydrophobic properties of water. Implicit solvent methodology comes at a price of number of approximations whose effects are often hard to estimate. Some familiar descriptors of molecular interaction, such as solute–solvent hydrogen bonds, are no longer explicitly present in the model; instead, they come in implicitly, in the mean-field way *via* a linear dielectric response, and contribute to the overall solvation energy. However, despite the fact that the methodology represents an approximation at a fundamental level, it has in many cases been successful in calculating various macromolecular properties[56-58]. Here we outline the implicit solvent methodologies implemented in AMBER.

Within the framework of the continuum model, a numerically exact way to compute the electrostatic potential $\phi(\boldsymbol{r})$ produced by molecular charge distribution $\rho_m(\boldsymbol{r})$, is based on the Poisson-Boltzmann (PB) approach in which the following equation (or its equivalent) must be solved; for simplicity we give its linearized form:

$$\nabla\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})=-4\pi\rho_m(\boldsymbol{r}) + \kappa^2\varepsilon(\boldsymbol{r})\phi(\boldsymbol{r}). \quad (1)$$

Here, $\varepsilon(\boldsymbol{r})$ represents the position-dependent dielectric constant which equals that of bulk water far away from the molecule, and is expected to decrease fairly rapidly across the solute/solvent boundary. The electrostatic screening effects of (monovalent) salt enter via the second term on the right-hand side of Eq. 1, where the Debye-Hückel screening parameter $\kappa \approx 0.1\text{Å}^{-1}$ at physiological conditions. Once the potential $\phi(\boldsymbol{r})$ is computed, the electrostatic part of the solvation free energy is given by:

$$\Delta G_{el} = \frac{1}{2}\sum_i q_i \left[\varphi(r_i) - \varphi(r_i)_{vac}\right] \quad (2)$$

where q_i are the partial atomic charges at positions $\boldsymbol{r}_i$ that make up the molecular charge density and $\phi(\boldsymbol{r}_i)|_{vac}$ is the electrostatic potential computed for the same charge distribution in the absence of the dielectric boundary, *e.g.* in vacuum. Full accounts of this theory are available elsewhere[59, 60].

The analytic generalized Born (GB) method is an approximate way to compute ΔG_{el}. The methodology has become popular, especially in MD applications, due to its relative simplicity and computational efficiency, compared to the more standard numerical solution of the PB equation[57, 61]. Within the GB models currently available in AMBER and NAB, each atom in a molecule is represented as a sphere of radius ρ_i with a charge q_i at its center; the interior of the atom is assumed to be filled uniformly with material of dielectric constant of 1. The molecule is surrounded by a solvent of a high dielectric ε_w (78.5 for water at 300 K). The GB model approximates ΔG_{el} by an analytical formula[19, 57],

$$\Delta G_{el} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)}\left(1-\frac{e^{-\kappa f^{GB}}}{\varepsilon_w}\right) \tag{3}$$

where r_{ij} is the distance between atoms i and j, R_i is the so-called *effective Born radii* of atom i, and f^{GB} is a certain smooth function of its arguments. The electrostatic screening effects of (monovalent) salt are incorporated [62] into Eq. 3 via the Debye-Hückel screening parameter $\kappa[\text{Å}^{-1}] \approx 0.316\sqrt{[salt][mol/L]}$.

A common choice[19] of f^{GB} is:

$$f^{GB} = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2 / 4R_i R_j)\right]^{1/2} \tag{4}$$

although other expressions have been tried[63-66]. The effective Born radius of an atom reflects the degree of its burial inside the molecule: for each charge, the "perfect" effective radius R_i satisfies:

$$\Delta G_i^{PB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon_w}\right)\frac{q_i^2}{R_i} \tag{5}$$

where ΔG_i^{PB} is the solvation energy (computed from a numerical solution to Eq. 1) for a single charge q_i in the dielectric environment of the full system. The effective radii depend on the molecule's conformation, and these need to be re-computed every time the conformation changes.

The efficiency of computing the effective radii is therefore a critical issue, and various approximations are normally made to accelerate the calculations. In particular, the so-called *Coulomb field approximation*, is

often used, which approximates the electric displacement around an atom by the Coulomb field $\bar{D}_i^0(\boldsymbol{r}) \equiv q_i \boldsymbol{r}/r^3$. Within this assumption, the following expression for R_i can be derived[57, 67]:

$$R_i^{-1} = \frac{1}{\rho_i} - \frac{1}{4\pi} \int_{solute} \theta(|r| - \rho_i) \frac{1}{r^4} d^3 r \tag{6}$$

where the integral is over the solute volume, excluding a sphere of radius ρ_i around atom *i*. Over the years, a number of slightly different intrinsic radii ρ_i have been proposed. A good set is expected to be transferable or to perform reasonably well in different types of problems. One example is the Bondi radii set originally proposed in the context of geometrical analysis of macromolecular structures, but later found to be useful in continuum electrostatics models as well[68].

For a realistic molecule, computing the integral in Eq. 6 is anything but trivial, so approximations are often made to obtain a closed-form analytical expression. The AMBER and NAB programs adopt the pairwise approach of Hawkins, Cramer and Truhlar[20, 69, 70], where the integral in Eq. 6 is approximated by a sum of terms over all other atoms ($j \neq i$) in the molecule:

$$R_i^{-1} = \frac{1}{\rho_i} + \sum_{j \neq i} f(d_{ij}, \rho_i, \rho_j) \tag{7}$$

Here d_{ij} is the distance between atoms *i* and *j*, and detailed expressions for the functions *f*() are given elsewhere[70, 71] These terms are exact for two overlapping spherical regions, and are approximate for points in space that are inside more than two atomic spheres. Although the scheme is approximate, it is completely analytical, so that forces can be computed as the derivatives of the energy, allowing standard applications of minimization and molecular dynamics to be carried out[21, 22].

1.3 AMBER Analysis

Another area of focus is to facilitate the analysis of molecular dynamics trajectories; this includes the ability to process, manipulate and analyze MD trajectory and coordinate files (`ptraj`) and also to extract (free) energetics (`mm-pbsa`). `ptraj` is a program that provides a set of "actions" that are performed sequentially on each snapshot of a set of trajectories, the format of which are auto-detected. Currently `ptraj` supports AMBER formats for trajectory and restart files, PDB, CHARMM binary trajectories of either endian (byte) order, and a simple binary

format. The "actions" that can be performed include best RMSd fitting, measurement of distances, angles, ring puckers, vectors, and atomic positional fluctuations, construction of radial distribution functions, calculation of mean-squared displacements (diffusion), 2D RMSd maps, and output of coordinate averaged structures. Trajectory manipulations supported include general periodic imaging, stripping atoms, translation, scaling, running averages, and saving the closest solvent molecules. Data that is accumulated during the trajectory processing (such as time series of scalars or vectors or matrices) is analyzed after the trajectories are processed sequentially with a series of "analyze" commands. More specialized features of the program include the hydration analysis (`grid`), hydrogen bond analysis (`hbond`), `matrix` facility, and clustering (`cluster`), each of which we describe in more detail below. If one is familiar with the C language, by following the detailed comments in the code it is relatively straightforward to add new actions that act on coordinate snapshots, or to add new analysis commands that work on the accumulated stack of scalars or vectors. Additionally, adding support for different coordinate file formats (for input or output) is also possible (albeit slightly less straightforward). The ease of extensibility of this code has led a number of different groups to extend the code.

Hbond: This facility provides a means to keep track of the time series, lifetime and occupancy of specific pair or triple distances and angles. To do this for all possible pair interactions—as a function of time—quickly becomes intractable (due to memory demands) so the set of interactions to consider is limited as specified by the user using the `donor` and `acceptor` commands (based on atom/residue name matching or general atom selection). Interactions between *electron pair* donors (single atoms) and *electron pair* acceptors (two atoms) are monitored (in contrast to *hydrogen* acceptors and donors). The following examples will keep track of all pair interactions between atoms named N3 in residues named GUA and the bond/angle formed to atoms N4 and H42 in residues named CYT.

```
donor GUA N3
acceptor CYT N4 H42
```

Multiple donor and acceptor commands may be specified. If the same atom name is specified twice in an acceptor interaction, the angle is ignored (i.e. only the distance is calculated); this is useful for monitoring pair interactions of single atoms to donors such as the interaction of ions (where one might specify "`acceptor CIO Na+ Na+`" to select atoms named Na+ in residues named CIO). In addition to specific pair interactions, a concept of general interactions has also been implemented to monitor solvent interactions. In this case, rather than keeping track of every separate solvent molecule and its interaction (which quickly

becomes memory intensive particularly if solvent-solvent interactions are monitored), the program only keeps track of a list (equal to the value of `solventneighbors` specified) of solvent molecules interacting with each donor or acceptor group. By default, six solvent interactions are stored, where solvent is defined arbitrarily by the user as a set of molecules (such as residues named WAT or residues named Na+). An example usage of the `hbond` command (post definition of the solvent and specification of the acceptor and donor lists) follows:

```
hbond series hb out hbond_wat.out \
  solventdonor WAT O \
  solventacceptor WAT O H1 \
  solventacceptor WAT O H2 \
  time 1.0 angle 120.0 distance 3.5
```

The keywords above turn on the time series (`series`) and name it `hb`, output a summary file called `hbond_wat.out`, setup the list of `solventdonor` and `solventacceptors`, specify that there is 1 ps between frames, set the angle cutoff such that the donor – acceptorH – acceptor angle is greater than 120.0 (although the angle stored internally and output is 180 minus this value) and that the donor – acceptor distance is less than 3.5 Å. A truncated version of the output (omitting standard deviations and more details) looks somewhat like the following:

```
 DONOR   ACCEPTOR    %occ  dist  angle  lifetime
| :5@N1 |:40@H3 :40@N3 | 99.93 2.955 15.50 1281.4 3159
| :40@O4 | :5@H61 :5@N6 | 98.26 2.975 17.90  60.3 517
| :39@O4 | :5@H61 :5@N6 |  1.86 3.296 50.73   1.1  5
| :5@O4' | solv accptr | 36.64 3.016 33.69   1.8  35
| :5@N7 | solv accptr | 132.69 2.958 26.33   3.0  67
| :5@N3 | solv accptr | 98.88 2.916 24.69   7.4 129
|solv dnr | :5@H62 :5@N6 | 93.75 3.085 29.61   3.0  48
```

The data shown is for some of the interactions with residue 5 of the DNA duplex with sequence d($CGA_4T_4CGA_4T_4CG$) from a ~27 ns MD simulation (in explicit solvent within a truncated octahedron unit cell applying the particle mesh Ewald method and standard MD simulation protocols [32]). Listed are the atoms involved in the hydrogen bond (or a designation that specific that the interaction is with generic solvent), the percent occupation, the average distance and angle (away from linear or 180.0°), the lifetime in picoseconds and the maximum number of continuous frames for which the interaction was present. For interaction with water molecules, the occupancy can be greater than 100% signifying that multiple waters are interacting. Planned extensions to the `hbond` facility include storing the distribution of distances (to avoid strict cutoffs on the hydrogen bond distance) and

enhancements that will facilitate dumping out and analysis of the individual time series. The `hbond` facility provides a means to track specific pair or triple interactions over the course of a simulation. The output shown above is augmented by an ascii representation of the time series of the interaction (where "darker" characters imply greater occupancy, i.e. " " for 0-5%, "." for 5-20%, "-" for 20-40%, "x" for 60-80%, "*" for 80-95% and "@" for 95-100% occupancy over equal spaced time intervals from the trajectory). In the first case, full occupancy is seen over the trajectory (representing one of the AT base pair hydrogen bonds which also transiently forms with the preceding thymine residue, and partially occupied water interactions.

```
| :40@O4  | :5@H61 :5@N6  |@@@@@@@@@@@@@@@@@@@@|
| :39@O4  | :5@H61 :5@N6  |          |
| :5@O4'  | solv accptr |---o--o--------o-o-o|
|solv dnr | :5@H62 :5@N6  |******@****@**@***@*|
```

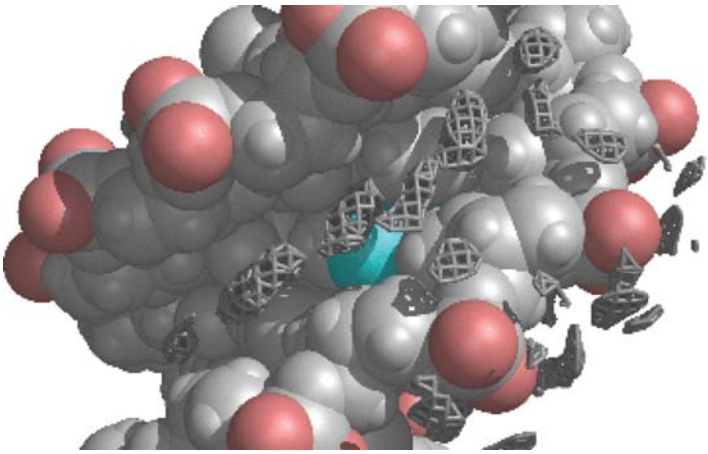

Figure 2-1. From ~27 ns of MD simulation of d($CGA_4T_4CGA_4T_4CG)_2$ the water oxygen density at ~3x (light gray) and ~ 4x bulk water (darker gray) is displayed (as a patchwork grid) on top of the straight coordinate averaged structure from 10-15ns with a view into the minor groove. The phosphate oxygens are highlighted (as a different shade) as is adenine residue 5 atom N3; the 100 x 100 x 100, 0.5 Å spacing grid, was centered on residues 4-6. In addition to the spine of hydration in the minor groove, hydration of the phosphates is evident.

Grid: The hydrogen bond analysis gives an indication of pair interactions but does not provide a visualization of how, for example, water interacts with a particular group. To look at solvation or ion density or effectively any atomic density around a particular group of atoms, a very simple grid procedure was developed that constructs an arbitrarily sized cubic grid and

counts the number of atoms (of a given type chosen by the user) within each grid element. The resulting grid can be output in X-PLOR density format for subsequent visualization by Chimera (see Figure 2-1), MIDAS, or VMD. In practice, prior to constructing the grid the region of interest needs to be centered, the trajectory fit to a common reference frame (usually the region interest), and the atoms imaged (as the current grid procedure does not image on the fly and therefore will put holes of density if atoms are in image cells). This method provides a simple and qualitative means to view regions of high density around the molecule of interest (such as in the grooves of DNA). To provide more quantitative estimates (less biased by motion of the region upon which the grid is centered and the chosen reference frame), radial distribution functions can be calculated (with the `radial` command).

Matrix/vector: With AMBER version 8.0, the vector facility of ptraj was greatly expanded and a new matrix facility developed by Holger Gohlke[25]. In addition, the analyze section of the code was greatly enhanced. It is now possible to follow the time correlation of arbitrarily defined vectors (or including vectors perpendicular to a least-squares fit plane) and to build distance, covariance, and related matrices which can subsequently be processed to estimate entropies, and to analyze and project estimated modes of vibration[25].

Cluster: A general purpose clustering library, implementing nine different clustering algorithms (hierarchical, single linkage, centroid linkage, complete linkage, K-means, centripetal, COBWEB, Bayesian, and self-organizing maps), has been integrated into the `ptraj` program and allows clustering of the trajectory snapshots based on pairwise RMSd or distance matrix comparisons. New trajectories (and/or average and/or representative structures) are output for each trajectory. To make clustering of large trajectories tractable, sieving has been implemented in a 2-pass approach such that every n^{th} frame is initially clustered then the initially skipped frames are added to the cluster that is most representative. Users can specify how many clusters are desired or set-up a pairwise distance metric cutoff value (representing the distance between clusters) to dynamically choose how many clusters should be formed. More detailed discussion of the performance is available elsewhere[72].

MM-PBSA: In addition to the general analysis provided by `ptraj`, `MM-PBSA` is a method developed to post-processes MD trajectory data to extract approximate free energies. The name—an abbreviation for molecular mechanics with Poisson-Boltzmann and surface area terms—came from collaboration between the authors (the Case and Kollman groups) and a test of the idea that performing simple energy analysis over a series of snapshots obtained from explicit solvent simulations and averaging of the results (to dampen the noise) might provide insight into the energetics[73]. Rather than

including the dominant and largely fluctuating water-water and water-solvent molecular mechanical energies, explicit water is stripped and replaced by an implicit model (PB or GB). This simple method provides a quick and easy means to understand binding free energies, the relative stability between two different conformations of the same molecule, or the effect of mutation (alanine scanning)[74]. This analysis is equivalent to the ES/IS method of Hermans[75, 76], similar to the LIE approaches of Aqvist[77, 78], and a direct extension (that involves averaging the results over an ensemble of configurations or coordinate snapshots) of well-known methods for estimating free energy[79].

$$G = < E_{solute} - TS + G_{solvation} > \quad (8)$$

As summarized in the equation above, the free energy is estimated as the average over a series of snapshots/configurations from the MD trajectory of the molecular mechanical energy of the solute (E_{solute}), an estimate of the entropy of the solute (TS) either from minimization and normal mode analysis or quasi-harmonic estimates from the covariance matrix, and the solvation free energy as calculated from an implicit solvent method such as Poisson-Boltzmann or generalized Born methods ($G_{solvation}$, potentially with the addition of a hydrophobic surface area term). All of these energy terms can be calculated within current versions of AMBER and this is done via a series of Perl language scripts supplied with the program. In application to nucleic acids, the methods have proven useful for estimating free energy differences between A- and B- DNA in water or water/ethanol solution[73], different sequences[80], different loop geometries[53, 81], and ligand interaction[82, 83], among other applications. The accuracy of the method is limited by the approximations. In general, this includes omitting all of the explicit solvent and ions. If a specific ion or water is important to mediate stability or a particular interaction, such molecules cannot be ignored as has been shown in studies of DNA minor groove binders[83] and also quadruplex DNA formation[84]. However, these molecules can be included as part of the solute. An additional limitation is the approximation of the entropy which can be noisy and difficult to converge in quasiharmonic approaches. This is most significant when attempting to calculate absolute entropies of binding and estimating the rotational and translation entropy loss upon binding. As discussed previously[84], estimates from both theory and experiment of the rotational and translational entropy loss upon binding of a drug-like molecule span a large range (from 3.0-30.0 kcal/mol). The standard harmonic approach to estimating the vibrational entropy (with the program `nmode` in AMBER) currently requires minimization in vacuo and this tends to significantly distort highly charged structures like nucleic acids. A promising

advance is the inclusion of second derivatives for the generalized Born model as is currently supported in NAB.

2. MOLECULAR DYNAMICS OF NUCLEIC ACIDS

As discussed in a recent review[30], this past decade may aptly be called the 10 ns era of nucleic acid MD simulation. An extensive set of simulation results for a wide variety of nucleic acid systems, on the 1-10 ns, have been published. We highlight some of these results and some unpublished work in what follows focusing on benchmarking the performance, reliability and accuracy of nucleic acid MD simulation.

After the initial successes in the 1994-1995 timeframe, applying Ewald methods to simulate nucleic acids in solution[9, 10, 85], a wide variety of simulations were initiated to assess the performance, complement experiment, and learn something new about sequence specific nucleic acid structure, dynamics, bending, backbone modification and unusual structures.

The initial studies investigated standard DNA and RNA duplexes and triplexes, including modified backbones, benchmarking the performance of the methods and investigating the results on different sequences, including various structures determined experimentally and poly-adenine tracts[12, 14, 17, 85-93]. These initial studies were able to show the differential structure and dynamics of DNA-RNA hybrids which adopted a mixed A/B-DNA structure compared to the flexible B-DNA and rigid A-RNA of the pure duplexes[12]. These studies also clearly demonstrated specific and differential ion association and hydration in the grooves depending on sequence and structure[12, 90]. A significant next step was to test whether the empirical force fields and the DNA in explicit solvent were sensitive to changes in the environment or surroundings. The obvious test was to investigate the A-B DNA transition which is caused by changes in relative humidity or water activity. Although some of the earlier force fields had shown unexpected spontaneous B-DNA to A-DNA transitions in explicit water[94], the Cornell et al. parm94.dat force field distributed with AMBER showed rapid A-DNA to B-DNA transitions on a nanosecond time scale[11] (and even faster transitions in implicit solvent[22]). Stabilizing the expected geometry was not sufficient to show the generality of the force field. To do this, it was necessary to demonstrate that A-DNA could be stabilized under conditions expected to stabilize A-DNA (such as in mixed water/ethanol, high salt or with tightly associated polyvalent ions) and ideally spontaneous B-DNA to A-DNA transitions could be observed. We were able to demonstrate stabilization of A-DNA in mixed water/ethanol[16, 95] and also spontaneous B-DNA to A-DNA transitions when hexaamminecobalt(III) ions were

bound[13]. Although we never succeeded with the Cornell et al. force field to see spontaneous B-DNA to A-DNA transitions under high monovalent salt conditions or in mixed water/ethanol (which is not unreasonable given the expected 10 μs or greater time scales for conformational transition[96, 97]), we were encouraged by the ability of the simple empirical force fields (lacking any explicit polarization) to represent these subtle effects of the surroundings. In an attempt to overcome these limitations, Langley iteratively optimized (based on extensive MD simulation) the BMS nucleic acid force field to facilitate rapid A-B transitions in small DNA duplex structures under the appropriate conditions [35]. Although this force field does appear to facilitate the A-B transitions, overall the structures sampled appear more rigid and crystal like than the softer geometries sampled by the Cornell et al. force field[98] and the time scales for the B to A transition may be too rapid. Work on better understanding the A-B equilibrium continues to date with various force fields[99, 100] and minimal hydration models[101, 102].

A next step towards validating the methods was the study of unusual nucleic acid structures such as DNA quadruplexes (as discussed in more detail in the Chapter by Spackova and co-workers), i-DNA and zipper structures and investigation of sequence dependent structure (bending, twisting) and dynamics. In general, the methods have proven incredibly robust and well able to handle virtually any type of nucleic acid structure in MD simulation. Emerging limitations, such as the convergence to incorrect loop geometries in G-DNA tetraplex structures[103], relate to force field limitations and the limited sampling afforded by computational limits which only allow routine simulation in the 10-50 ns time scales.

In addition to correct representation of structure, the MD simulations have also allowed more detailed understanding of sequence and structure specific dynamics. In our early simulations of DNA-RNA hybrids, we showed that A-form geometries are relatively rigid compared to B-form duplexes, while mixed A/B geometries as seen in the hybrids have a flexibility between pure A- and B-form duplexes[12]. A more detailed analysis of sequence specific dynamics, demonstrated trends consistent with experiment, excepting that the Cornell et al. force field is slightly more flexible than expected[98, 104]. We have also demonstrated that the structural fluctuations also effect the electronic structure using DFT calculations to analyze snapshots extracted from MD simulation[105, 106]. The sequence specific structure and dynamics have also been investigated in a large scale collaborative effort (the ABC consortium) of a number of different research groups; two papers have resulted that highlight the performance of the methods[18, 107].

Although water and ions are integral to nucleic acid structure, and most of the studies discussed so far in this chapter included explicit ions and

counterions, there has been considerable progress in the development of implicit solvent models (as mentioned) and they have proven worthwhile in nucleic acid simulation. The first molecular dynamics simulations carried out using the AMBER generalized Born (GB) codes were on duplex DNA[22], and our early experiences with this solvation model have been reviewed[21]. In general, the basic goal of a continuum solvent model, namely that the results should closely mimic those using explicit solvent models, are achieved, at least for stable/regular helical structures of DNA and RNA. As an additional example, we discuss here both unpublished GB and explicit solvent simulations of the B-DNA duplex d(GTGACTGACTGACTG)-d(CAGTCA GTCAGTCAC). The extended (up to 80 ns) explicit solvent simulations were carried out as a part of an extensive study of sequence-dependence effects in DNA duplexes, whose results are reported elsewhere [18, 107]. Since the reported results (on the 15 ns time scale) have already been extensively characterized, comparisons to 200+ ns simulations with the GB model can focus on differences between the implicit and explicit solvent models. The GB calculations described here used the *igb = 2* model in AMBER (originally applied to proteins)[108], setting a salt concentration to 0.1 M, and using a nonbonded cutoff of 20 A. Langevin dynamics were used, with a target temperature of 300K and a friction constant of 5 ps^{-1}. This friction constant is less than that expected for water, but is one that is appropriate for rapid exploration of configuration space[109]. Hence, one should not expect the time-dependence of the GB results to have physical significance, but the configurations explored should reflect those preferred by our force field.

As we have found in our earlier DNA simulations, the basic double-helical character remains nearly the same in explicit and implicit solvent simulations, at least for 1-10 nanoseconds of simulation, and that over the course of longer simulations the all-atom profiles of the RMSd versus time are similar. During the MD, the structures quickly move to a metastable set of structures on the range of 6-7 Å (measured over all 15 base pairs) from the starting structure. Figure 2-2 shows the time-dependence of the all-atom mass-weighted RMSd difference between the starting structure (which is an idealized B-form helix) and snapshots taken at ps time intervals during the simulation over the first 80 ns (with the entire GB simulation out to over 200 ns shown on the inset).

A comparison of the average all-atom RMSd over the 20-40 ns interval to the starting structure is 5.6+/–0.8 Å for the explicit water simulation and 6.7+/–0.7 Å for the GB simulation. Although these RMSd differences may appear alarmingly large, the main differences between the sampled configurations and the starting structures are primarily related to the slightly smaller twist at each basepair step in the simulation compared to an idealized form, a slight tendency to roll into the major groove (which shifts the RMSd

closer towards an A-form like helix despite maintaining B-form sugar puckers and helicoidal parameters), some terminal group fraying in the GB simulations, and population of unexpected α, γ backbone sub-states; these preferences for lower twist (which arise from the use of the Cornell et al., AMBER 94 force field)[4] have been discussed in detail elsewhere[11, 32, 110, 111].

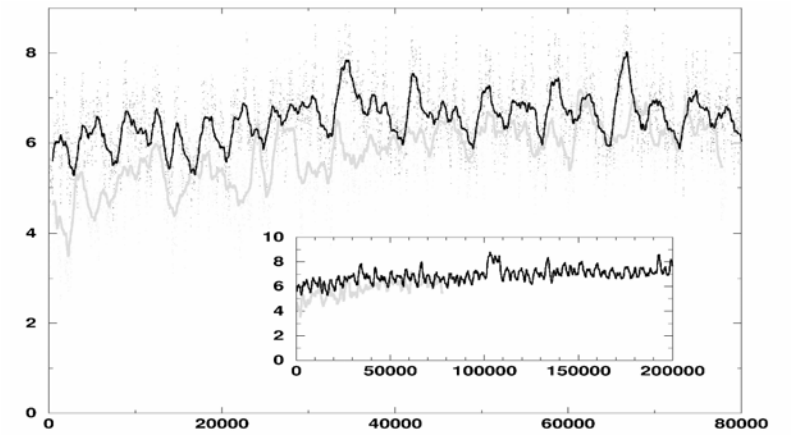

Figure 2-2. All-atom mass-weighted RMSd (Å) vs. time (ps) of the MD snapshots at ps intervals compared to the idealized canonical B-DNA starting model. The explicit water simulation is shown in gray and the GB simulation in black. In both plots, a running average has been applied over 1000 ps to smooth the data for easier visualization with the raw data at 10 ps intervals in the outer plot shown as dots.

If instead the RMSd is focused on the central 5 base pairs, the average all-atom RMSd over the first 40 ns for the explicit solvent simulation is 2.6+/–0.4 Å, and 2.9+/–0.4 Å for the GB simulations.

Additionally, the average structures for the two simulations (GB and explicit solvent) are rather close (1.78 Å, all-atom, for the 9-10 ns straight coordinate averaged structures) until later time frames in the GB simulation where (sometimes reversible) fraying of the helix ends is observed. This is shown in Figure 2-3 with snapshots taken after 13, 14 and 15 ns of the GB MD simulation. All of the structures in the first 13 ns resemble the one at the left, with all Watson-Crick hydrogen bonds intact and a stable helix.

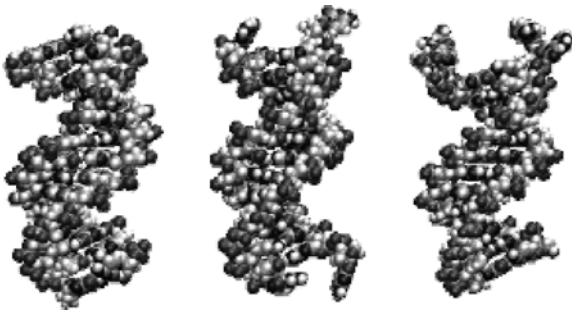

Figure 2-3. Snapshots of the simulated DNA duplex structure from the GB simulations at 13, 14 and 15 nanoseconds.

Starting at about 10 ns in this simulation, some fraying of the end groups becomes apparent, as shown in the middle part of the figure above. This is not unexpected, although the correct extent of end-fraying in 15-mer DNA duplexes is not known. In some cases, as at the bottom of the right-hand view in the figure, the Watson-Crick bonds re-form, but in other cases the end fraying is irreversible on the time scales investigated here and as the simulation is extended beyond 100 ns, end-group fraying is seen at both ends of the helix with the frayed bases tucking back into the grooves. Complicating the discussion is the existence of transitions about the backbone α and γ torsion angles, which undergo a "crankshaft" motion that appears to be irreversible on the time scales sampled here in both the implicit and explicit solvent simulations.

Because of these problems with the long-term behavior of the helix, and since the initial ABC simulations of this sequence were analyzed in detail on the 15 ns time scale, we have chosen here to use the middle portion of the helix for the first 15 ns to analyze internal fluctuations. One set of parameters of great interest to the interpretation of NMR data are order parameters for C-H bonds, which are computed from time-correlation functions[112]:

$$S^2 = \lim_{t\to\infty} < P_2(\cos\theta(\tau)) > \qquad (9)$$

Here $\theta(\tau)$ is the angle between the direction of the C-H vector at time 0 and its direction at time τ, and P_2 is a second-order Legendre polynomial. The brackets indicate an average over all members of an equilibrium ensemble, and the time delay is long enough to include all internal motions; overall rotational tumbling is assumed to be an independent motion, and has been removed from the analysis. These order parameters are convenient measures

of the extent of internal fluctuations that can be determined from heteronuclear NMR measurements, and also be readily determined from molecular dynamics simulations. They thus form a common ground for comparisons of MD simulations to NMR data[113].

Figure 2-4 shows some characteristic behavior for these sorts of time correlation functions, using the C8-H8 vector of two adenine bases for examples. At zero time, the angle θ is zero, and the correlation function must go to unity.

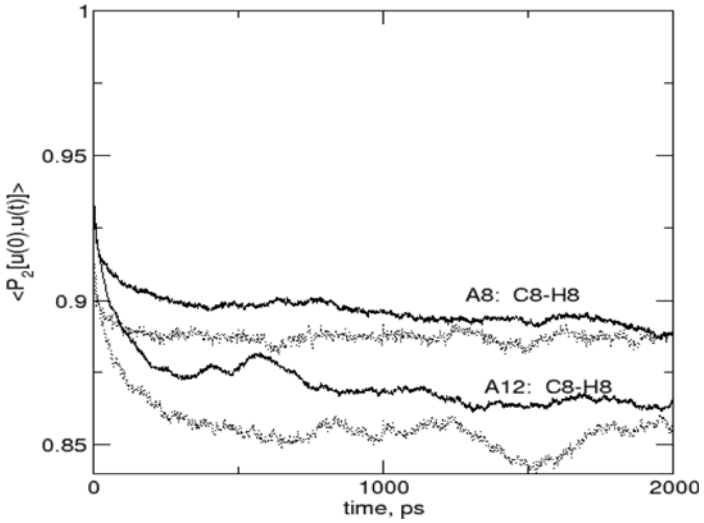

Figure 2-4. Time correlation functions for the C8-H8 inter-nuclear vector in adenine residue 8 (A8), and the corresponding vector in adenine residue 12 (A12). Solid lines show results from the explicit solvent simulations, dotted lines are from the GB results.

At very short times, not visible in this picture, the angle changes due to fast (sub-picosecond) vibrational motion; after this, there may be additional decays of the correlation function arising from larger scale conformational changes. In most cases, as seen here, there is an approximate plateau value that is usually achieved in the first 1-2 nanoseconds of motion. This plateau gives the value of S^2. The examples in Figure 2-4 are representative ones: values of the plateau region are clearly evident in an approximate fashion, but noise in the simulations limits the precision to which they can be determined. Nevertheless, it is clear that the values reached are smaller for A12 than for A8, and that explicit and implicit results are roughly the same. As expected, the rate at which these plateau values are reached is clearly

longer for the explicit solvent simulation, although even in the generalized Born calculation it can take 0.5 ns for the fluctuations to build up.

More of an overall view of the amplitudes of these motions are shown in Figures 2-5 and 2-6, which plot the order parameters for bases and sugars for the explicit and implicit solvent models. The extent of base motion is quite similar in the two simulations, with high order parameters (less motion) in explicit solvent being very strongly predictive of a high order parameter in the GB simulation. The situation is a little more complex for the C1'-H1' internuclear vector in the sugars. The general trend in the two simulations is about the same, with more motion (as expected) near the ends of each strand, and less motion in the middle. But here the GB simulation is more smooth as a function of position within the helix, whereas the explicit solvent results show more structure, with purine sugars tending to be more mobile than the pyrimidines. Further studies will be required to see if this is a general trend, and to identify what differences in the solvent models might lead to this sort of behavior.

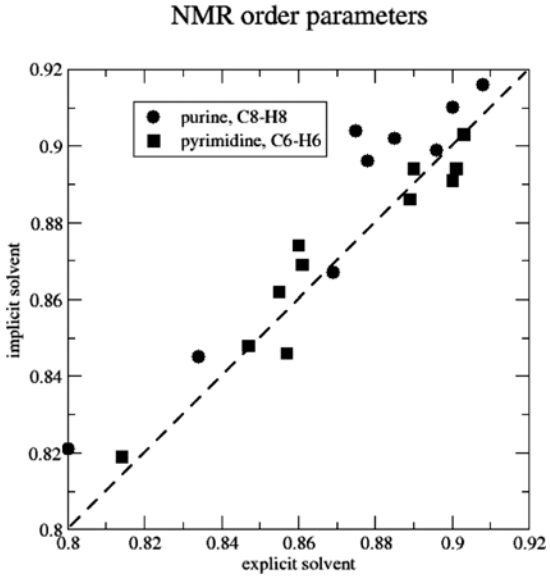

Figure 2-5. Comparison of order parameters computed for C-H vectors purine and pyrimidine bases, comparing results for the explicit and implicit solvent simulations.

Overall, generalized Born solvation models are still quite new, and development continues that have them be a more faithful description of

explicit solvent results. The examples shown here illustrate that the overall behavior for nucleic acids is acceptable (so that these force fields are attractive models for X-ray or NMR refinement studies, for example[114, 115]) but that one must still treat details of the simulations as suspect.

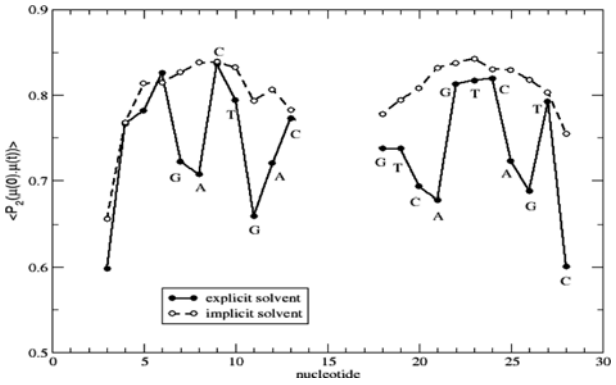

Figure 2-6. Order parameters for the C1'-H1' vector in sugar residues, comparing explicit solvent (solid lines) with generalized Born results (dashed lines).

3. CHALLENGES FOR THE FUTURE

A focusing and mystifying question that readily arises when studying nucleic acids is: *what is the correct answer*? This is a broad question that relates not only to the accuracy and reliability of the simulation results, but also to experiment and our incomplete understanding of nucleic acid structure, dynamics and interactions. Experiment is sometimes inconclusive and potentially influenced by artifacts (due to packing, the surroundings, end-effects, etc); moreover, the structures of nucleic acids are profoundly sensitive to their environment. This sensitivity provides challenge for the methods. Despite this caveat, in the 1-10 ns time scale, simulation has given tremendous insight into ligand interaction, unusual nucleic acid structures, backbone modification and general sequence specific structure and dynamics. However, as we push the simulation methods to longer size and time scales, key limitations become more readily apparent, including the need to obtain more data from experiment to better verify the simulation methods. Key questions whose answers will aid in the optimization of the nucleic acid force fields include:

- What is the rate of sugar repuckering?
- What is the rate (time scale) and nature of the A-B transition for a small DNA duplex in solution?
- What is the rate of B_I to B_{II} or α/γ crankshaft backbone transitions?
- What sub-states of the sugar ring and backbone should be populated and to what extent?
- Can additive models properly represent ionic effects (ionic strength, identify) and specific ion/solvent interaction?
- Do ions cause DNA to bend and how long do ions and solvent associate with nucleic acids?

The lack of complete understanding makes nucleic acids a fun and challenging system for study and justifies the application of theoretical methods in an attempt to better understand sequence dependent structure, dynamics and recognition. To date we have had clear successes. Examples include A–>B transitions, drug-nucleic acid interaction, and accurate representation of sequence specific dynamics and structural differences. On the other hand, the force fields all do not agree regarding repuckering rates or base pair opening time scales/energetics, or backbone sub-state populations. Moreover, as we push to longer time and size scales, hidden problems may emerge and sampling limitations may hinder understanding. An example is the unexpectedly high population of α,γ:g +,t states in DNA duplex simulation. These issues provide impetus to improve the simulation methods and force fields applied to nucleic acids.

ACKNOWLEDGEMENTS

We wish to thank all the AMBER developers over the years for their efforts and especially acknowledge the vision and dynamic of Peter Kollman. TEC would like to point out that he never would have continued on a life-long path of nucleic acid studies if it were not for the first few days with the PME method and A-B transitions that so excited Peter such that his enthusiasm did inspire and still infects me! An additional note of thanks goes to Tom Darden (NIEHS) for his tremendous efforts in getting particle mesh Ewald methods into AMBER and the ABC consortium of DNA simulators for working together to better understand applications of molecular dynamics methods to nucleic acids. TEC would like to acknowledge support from NSF CHE-0326027 and generous grants of computer time from the LRAC (large resources allocations committee) MCA01S027 at the Pittsburgh Super-computing Center, the National Center for Supercomputing Applications at the University of Illinois, and the San Diego Supercomputing Center, and also

significant resources on the Arches metacluster at the Center for High Performance Computing at the University of Utah (NIH-1S10RR017214-01). DAC was partially supported by NIH grant GM-57513.

REFERENCES

1. Seibel, G.L., U.C. Singh, and P.A. Kollman (1985). A molecular dynamics simulation of double-helical B-DNA including counterions and water. *Proc. Nat. Acad. Sci.* **82**, 6537-6540.
2. Sanbonmatsu, K.Y. and S. Joseph (2003). Understanding discrimination by the ribosome: Stability testing and goove measurement of codon-anticodon pairs. *J. Mol. Biol.* **328**, 33-47.
3. Tung, C.S. and K.Y. Sanbonmatsu (2004). Atomic model of the Thermus thermophilus 70S ribosome developed in silico. *Biophys J.* **87**, 2714-2722.
4. Cornell, W.D., *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.* **117**, 5179-5197.
5. Cheatham, T.E., III, P. Cieplak, and P.A. Kollman (1999). A modified version of the Cornell *et al*. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **16**, 845-862.
6. Wang, J., P. Cieplak, and P.A. Kollman (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* **21**, 1049-1074.
7. Darden, T.A., D.M. York, and L.G. Pedersen (1993). Particle mesh Ewald - An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092.
8. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen (1995). A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **103**, 8577-8593.
9. Cheatham, T.E., III, J.L. Miller, T. Fox, T.A. Darden, and P.A. Kollman (1995). Molecular dynamics simulations on solvated biomolecular systems - The particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. *J. Amer. Chem. Soc.* **117**, 4193-4194.
10. York, D.M., W. Yang, H. Lee, T.A. Darden, and L. Pedersen (1995). Toward accurate modeling of DNA: the importance of long-range electrostatics. *J. Amer. Chem. Soc.* **117**, 5001-5002.
11. Cheatham, T.E., III and P.A. Kollman (1996). Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259**, 434-444.
12. Cheatham, T.E., III. and P.A. Kollman (1997). Molecular dynamics simulations highlight the structural differences in DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Amer. Chem. Soc.* **119**, 4805-4825.
13. Cheatham, T.E., III and P.A. Kollman (1997). Insight into the stabilization of A-DNA by specific ion association: Spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaammine cobalt(III). *Structure* **5**, 1297-1311.

14. Young, M.A., G. Ravishanker, and D.L. Beveridge (1997). A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions and solvation. *Biophys. J.* **73**, 2313-2336.
15. Cheatham, T.E., III and P.A. Kollman (1998). *Molecular dynamics simulation of nucleic acids in solution: How sensitive are the results to small perturbations in the force field and environment*, in *Structure, motion, interactions and expression of biological macromolecules*, M. Sarma and R. Sarma, Editors, Adenine Press: Schenectady, NY. p. 99-116.
16. Sprous, D., M.A. Young, and D.L. Beveridge (1998). Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: Theoretical considerations of the A–>B transition. *J. Phys. Chem. B.* **102**, 4658-4667.
17. Young, M.A. and D.L. Beveridge (1998). Molecular dynamics simulations of an oligonucleotide duplex with adenine-tracts phased by a full helix turn. *J. Mol. Biol.* **281**, 675-687.
18. Beveridge, D.L., *et al.* (2004). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **87**, 3799-3813.
19. Still, W.C., A. Tempczyk, R.C. Hawley, and T. Hendrickson (1990). Semi analytical treatment of solvation for molecular mechanics and dynamics. *J. Amer. Chem. Soc.* **112**, 6127-6128.
20. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**, 122-129.
21. Tsui, V. and D.A. Case (2001). Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopol.* **56**, 275-291.
22. Tsui, V. and D.A. Case (2000). Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Amer. Chem. Soc.* **122**, 2489-2498.
23. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D. States, J,S. Swaminathan, and M. Karplus (1983). CHARMm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.
24. Pearlman, D.A., D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structure and energetic properties of molecules. *Comp. Phys. Comm.* **91**, 1-41.
25. Case, D.A., *et al.* (2005). The AMBER biomolecular simulation programs. *J. Comp. Chem.* **26**, 1668-1688.
26. Beveridge, D.L. and K.J. McConnell (2000). Nucleic acids: theory and computer simulation, Y2K. *Cur. Op. Struct. Biol.* **10**, 182-196.
27. Cheatham, T.E., III and P.A. Kollman (2000). Molecular dynamics simulation of nucleic acids. *Ann. Rev. Phys. Chem.* **51**, 435-471.
28. Orozco, M., A. Perez, A. Noy, and F. Javier Luque (2003). Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **32**, 350-364.
29. Cheatham, T.E., III (2004). Simulation and modeling of nucleic acid structure, dynamics and interactions. *Cur. Opin. Struct. Biol.* **14**, 360-367.

30. Cheatham, T.E., III (2005). Molecular modeling and atomistic simulation of nucleic acids. *Ann. Reports Comp. Chem.* **1**, 75-89.
31. Norberg, J. and L. Nilsson (2002). Molecular dynamics applied to nucleic acids. *Acc. Chem. Res.* **35**, 465-472.
32. Cheatham, T.E., III and M.A. Young (2001). Molecular dynamics simulations of nucleic acids: Successes, limitations and promise. *Biopolymers* **56**, 232-256.
33. MacKerell, A.D., Jr. (2004). Empirical force fields for biological macromolecules: Overview and issues. *J. Comp. Chem.* **25**, 1584-1604.
34. Reddy, S.Y., F. Leclerc, and M. Karplus (2003). DNA polymorphism: A comparison of force fields for nucleic acids. *Biophys. J.* **84**, 1421-1449.
35. Langley, D.R. (1998). Molecular dynamics simulations of environment and sequence dependent DNA conformation: The development of the BMS nucleic acid force field and comparison with experimental results. *J. Biomol. Struct. Dyn.* **16**, 487-509.
36. Soares, T.A., P.H. Hunenberger, M.A. Kastenholz, V. Krautler, T. Lenz, R.D. Lins, C. Oostenbrink, and W.F. van Gunsteren (2005). An improved nucleic acid parameter set for the GROMOS force field. *J. Comp. Chem.* **26**, 727-737.
37. Cheatham, T.E., III, B.R. Brooks, and P.A. Kollman (2000). *Molecular modeling of nucleic acid structure*, in *Current Protocols in Nucleic Acid Chemistry*, S.L. Beaucage, et al., Editors, Wiley: New York. p. 7.5.1-7.5.12.
38. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case (2004). Development and testing of a general amber force field. *J. Comp. Chem.* **25**, 1157-1174.
39. Jakalian, A., B.L. Bush, D.B. Jack, and C.I. Bayly (2000). Fast, efficient generation of high quality atomic charges. AM1-BCC model: I. method. *J. Comp. Chem.* **21**, 132-146.
40. Jakalian, A., D.B. Jack, and C.I. Bayly (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623-41.
41. Macke, T.J. and D.A. Case, *NAB (Nucleic Acid Builder)*. 1996, The Scripps Research Institute: La Jolla, CA.
42. Macke, T. and D.A. Case (1998). *Modeling unusual nucleic acid structures*, in *Molecular modeling of nucleic acids*, N.B. Leontis and J. Santa Lucia, Editors, ACS: Washington. p. 379-393.
43. Lu, X.J. and W.K. Olson (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**, 5108-21.
44. Mueller, F. and R. Brimacombe (1997). A new model for the three-dimensional folding of Escherichia coli 16 S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 A. *J. Mol. Biol.* **271**, 524-44.
45. Burks, J., C. Zwieb, F. Muller, I. Wower, and J. Wower (2005). Comparative 3-D Modeling of tmRNA. *BMC Mol. Biol.* **6**, 14.
46. Massire, C. and E. Westhof (1998). MANIP: an interactive tool for modelling RNA. *J. Mol. Graph. Model.* **16**, 197-205, 255-7.
47. Major, F., D. Gautheret, and R. Cedergren (1993). Reproducing the three-dimensional structure of a tRNA molecule from structural constraints. *Proc. Natl. Acad. Sci.* **90**, 9408-12.

48. Tung, C.S. and E.S. Carter, 2nd (1994). Nucleic acid modeling tool (NAMOT): an interactive graphic tool for modeling nucleic acid structures. *Comput. Appl. Biosci.* **10**, 427-33.
49. Diaz, N., D. Suarez, T.L. Sordo, and K.M. Merz, Jr. (2001). A theoretical study of the aminolysis reaction of lysine 199 of human serum albumin with benzylpenicillin: consequences for immunochemistry of penicillins. *J. Am. Chem. Soc.* **123**, 7574-83.
50. Park, H., E.N. Brothers, and K.M. Merz, Jr. (2005). Hybrid QM/MM and DFT investigations of the catalytic mechanism and inhibition of the dinuclear zinc metallo-b-lactamase CcrA from Bacteroides fragilis. *J. Am. Chem. Soc.* **127**, 4232-41.
51. Wu, X. and B.R. Brooks (2005). Isotropic periodic sum: a method for the calculation of long-range interactions. *J. Chem. Phys.* **122**, 44107.
52. Roitberg, A. and R. Elber (1991). Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* **95**, 9277-9286.
53. Simmerling, C., J.L. Miller, and P.A. Kollman (1998). Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a non-helical nucleic acid. *J. Amer. Chem. Soc.* **120**, 7149-7155.
54. Feig, M., J. Karanicolas, and C.L. Brooks, 3rd (2004). MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph Model* **22**, 377-95.
55. Sagui, C. and T.A. Darden (1999). Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Ann. Rev. Biophys. Biomol. Struct.* **28**, 155-179.
56. Cramer, C.J. and D.G. Truhlar (1999). Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem Rev* **99**, 2161-2200.
57. Bashford, D. and D.A. Case (2000). Generalized Born models of molecular solvation effects. *Ann. Rev. Phys. Chem.* **51**, 129-152.
58. Feig, M. and C.L.I. Brooks (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14**, 217-224.
59. Simonson, T. (2003). Electrostatics and dynamics of proteins. *Rep. Prog. Phys.* **66**, 737-787.
60. Baker, N.A. (2004). Poisson-Boltzmann methods for biomolecular electrostatics. *Meth. Enzym.* **383**, 94-118.
61. Feig, M., W. Im, and C.L. Brooks, 3rd (2004). Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* **120**, 903-11.
62. Srinivasan, J., M.W. Trevathan, P. Beroza, and D.A. Case (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426-434.
63. Jayaram, B., Y. Liu, and D.L. Beveridge (1998). A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.* **109**, 1495-1471.
64. Onufriev, A., D.A. Case, and D. Bashford (2002). Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comp. Chem.* **23**, 1297-1304.
65. Grycuk, T. (2003). Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *J. Comp. Chem.* **119**, 4817-4826.

66. Sigalov, G., P. Scheffel, and A. Onufriev (2005). Incorporating variable dielectric environments into the generalized Born model. *J. Chem. Phys.* **122**, 094511.
67. Onufriev, A., D. Bashford, and D.A. Case (2000). Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B.* **104**, 3712-3720.
68. Bondi, A. (1964). van der Waals volumes and radii. *J. Phys. Chem.* **68**, 441-451.
69. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar (1997). Parameterized model for aqueous free energies of solvation using geometry-dependent atomic surface tensions with implicit electrostatics. *J. Phys. Chem. B.* **101**, 7147-7157.
70. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar (1996). Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824-19839.
71. Schaefer, M. and C. Froemmel (1990). A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J Mol Biol* **216**, 1045-66.
72. Tanner, S.W., N. Thompson, and T.E. Cheatham, III (2005). Cluster analysis of molecular dynamics trajectories: Insight into nucleic acid structure, dynamics and interactions. *[submitted]*.
73. Srinivasan, J., T.E. Cheatham, III, P. Cieplak, P.A. Kollman, and D.A. Case (1998). Continuum solvent studies of the stability of DNA, RNA and phosphoramidate helices. *J. Amer. Chem. Soc.* **120**, 9401-9409.
74. Kollman, P.A., *et al.* (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889-897.
75. Vorobjev, Y.N., J.C. Almagro, and J. Hermans (1998). Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins.* **32**, 399-413.
76. Vorobjev, Y.N. and J. Hermans (1999). ES/IS: Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* **78**, 195-205.
77. Froloff, N., A. Windemuth, and B. Honig (1997). On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Prot. Sci.* **6**, 1293-1301.
78. Hansson, T., J. Marelius, and J. Aqvist (1998). Ligand binding affinity prediction by linear interaction energy methods. *J. Comp. Aided Mol. Des.* **12**, 27-35.
79. Misra, V.K., K.A. Sharp, R.A. Friedman, and B. Honig (1994). Salt effects on ligand-DNA binding. Minor groove binding antibiotics. *J. Mol. Biol.* **238**, 245-263.
80. Cheatham, T.E., III, J. Srinivasan, D.A. Case, and P.A. Kollman (1998). Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.* **16**, 265-280.
81. Srinivasan, J., J.L. Miller, P.A. Kollman, and D.A. Case (1998). Continuum solvent studies of the stanbility of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **16**, 671-682.
82. Gouda, H., I.D. Kuntz, D.A. Case, and P.A. Kollman (2003). Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopol.* **68**, 16-34.

83. Spackova, N., T.E. Cheatham, III, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza, and J. Sponer (2003). Molecular dynamics simulations and thermodynamic analysis of DNA-drug complexes. Minor groove binding between 4',6-diamidino-2-phenylindole (DAPI) and DNA duplexes in solution. *J. Amer. Chem. Soc.* **125**, 1759-1769.
84. Stefl, R., T.E. Cheatham, III, N. Spackova, E. Fadrna, I. Berger, J. Koca, and J. Sponer (2003). Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamical analysis of the substates. *Biophys. J.* **85**, 1787-1804.
85. Cheng, Y.-K. and B.M. Pettitt (1995). Solvent effects on model d(CG-G)$_7$ and d(TA-T)$_7$ DNA triplex helices. *Biopol.* **35**, 457-473.
86. Young, M.A., J. Srinivasan, I. Goljer, S. Kumar, D.L. Beveridge, and P.H. Bolton (1995). Structure determination and analysis of local bending in an A-tract DNA duplex: Comparison of results from crystallography, nuclear magnetic resonance, and molecular dynamics simulation on d(CGCAAAAATGCG). *Methods in Enzymology* **261**, 121-144.
87. Norberg, J. and L. Nilsson (1996). Constant pressure molecular dynamics simulations of the dodecamers: d(GCGCGCGCGCGC)2 and r(GCGCGCGCGCGC)2. *J. Chem. Phys.* **104**, 6052-6057.
88. Shields, G.C., C.A. Laughton, and M. Orozco (1997). Molecular dynamics simulations of the d(TAT) triplex helix. *J. Amer. Chem. Soc.* **119**, 7463-7469.
89. Flatters, D., M. Young, D.L. Beveridge, and R. Lavery (1997). Conformational properties of the TATA-box binding sequence of DNA. *J. Biomol. Struct. Dyn.* **14**, 757-765.
90. Young, M.A., B. Jayaram, and D.L. Beveridge (1997). Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets. *J. Amer. Chem. Soc.* **119**, 59-69.
91. Spector, T.I., T.E. Cheatham, III, and P.A. Kollman (1997). Unrestrained molecular dynamics of photodamage DNA in aqueous solution. *J. Amer. Chem. Soc.* **119**, 7095-7104.
92. Luo, J. and T.C. Bruice (1998). Nanosecond molecular dynamics of hybrid triplex and duplex of polycation deoxyribonucleic guanidine strands with a complementary DNA strand. *J. Amer. Chem. Soc.* **120**, 1115-1123.
93. Shields, G.C., C.A. Laughton, and M. Orozco (1998). Molecular dynamics simulation of a PNA-DNA-PNA triple helix in aqueous solution. *J. Amer. Chem. Soc.* **120**, 5895-5904.
94. Yang, L.Q. and B.M. Pettitt (1996). B to A transition of DNA on the nanosecond time scale. *J. Phys. Chem.* **100**, 2564-2566.
95. Cheatham, T.E., III, M.F. Crowley, T. Fox, and P.A. Kollman (1997). A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci.* **94**, 9626-9630.
96. Jose, D. and D. Porschke (2005). The Dynamics of the B-A transition of natural DNA double helices. *J. Am. Chem. Soc.* **127**, 16120-16128.
97. Jose, D. and D. Porschke (2004). Dynamics of the B-A transition of DNA double helices. *Nuc. Acids Res.* **32**, 2251-8.
98. Lankas, F., J. Sponer, J. Langowski, and T.E. Cheatham, III (2003). DNA base-pair step deformability inferred from molecular dynamics simulation. *Biophys. J.* **85**, 2872-2883.
99. Pastor, N. (2005). The B- to A-DNA transition and the reorganization of solvent at the DNA surface. *Biophys. J.* **88**, 3262-75.

100. Banavali, N.K. and B. Roux (2005). Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.* **127**, 6866-76.
101. Mazur, A.K. (2002). DNA dynamics in a water drop without counterions. *J. Am. Chem. Soc* **124**, 14707-15.
102. Mazur, A.K. (2003). Titration *in silico* of reversible B<->A transitions in DNA. *J. Amer. Chem. Soc.* **125**, 7849-7859.
103. Fadrna, E., N. Spackova, R. Stefl, J. Koca, T.E. Cheatham, III, and J. Sponer (2004). Molecular dynamics simulations of guanine quadruplex loops: Advances and force field limitations. *Biophys. J.* **87**, 227-242.
104. Lankas, F., J. Sponer, J. Langowski, and T.E.I. Cheatham (2004). DNA deformability at the base pair level. *J. Amer. Chem. Soc.* **126**, 4124-4125.
105. Lewis, J.P., J. Pikus, T.E. Cheatham, III, E.B. Starikov, H. Wang, J. Tomfohr, and O.F. Sankey (2002). A comparison of electronic states in periodic and aperiodic poly(dA)-poly(dT) DNA. *Phys. Stat. Sol. (b)* **233**, 90-100.
106. Lewis, J.P., T.E. Cheatham, III, H. Wang, E.B. Starikov, and O.F. Sankey (2003). Dynamically amorphous character of electronic states in poly(dA)-poly(dT) DNA. *J. Phys. Chem. B* **107**, 2581-2587.
107. Dixit, S.B*., et al.* (2005). Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J*, [in press].
108. Onufriev, A., D. Bashford, and D.A. Case (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383-94.
109. Loncharich, R.J., B.R. Brooks, and R.W. Pastor (1992). Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **32**, 523-35.
110. Feig, M. and B.M. Pettitt (1997). Experiment vs force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem.* **101**, 7361-7363.
111. Real, A.N. and R.J. Greenall (2000). B->A->B transitions in a molecular dynamics trajectory of low salt DNA solution. *J. Mol. Model.* **6**, 654-658.
112. Lipari, G. and A. Szabo (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559.
113. Case, D.A. (2002). Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* **35**, 325-31.
114. Xia, B., V. Tsui, D.A. Case, H.J. Dyson, and P.E. Wright (2002). Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR* **22**, 317-31.
115. Moulinier, L., D.A. Case, and T. Simonson (2003). Reintroducing electrostatics into protein X-ray structure refinement: bulk solvent treated as a dielectric continuum. *Acta Crystallogr D. Biol Crystallogr* **59**, 2094-103.