

Molecular Modeling of Nucleic Acid Structure: Setup and Analysis

ISSUES IN THE SIMULATION OF NUCLEIC ACIDS

From the information presented in previous units (UNITS 7.5, 7.8 & 7.9), one should have a reasonable understanding of the various trade-offs that are necessary to model nucleic acid structures. For instance, when choosing an energy representation and a means to sample relevant conformations, there is a trade-off between detail, sampling, and computational cost. Although the discussions thus far have presented basic means for modeling nucleic acids, some important details have not been sufficiently addressed. Reasonable questions that remain are: (1) what empirical molecular mechanical force field is appropriate; (2) if one wants to run an accurate simulation of a nucleic acid in an explicit solvent, how does one set up and equilibrate the system; and (3) how does one analyze molecular dynamics trajectories? This unit provides the answers to some of these questions and outlines a protocol for accurate simulation of nucleic acids.

Which Molecular Mechanics Force Field Is Appropriate?

This is a difficult question. The answer is complicated, often contentious, and in part depends on what representation (i.e., implicit versus explicit solvent, internal coordinate versus all-atom) is applied. With an internal coordinate representation, where bonds and angles are fixed, the most widely used molecular mechanical force field for nucleic acids is the FLEX force field within JUMNA (Lavery et al., 1995). Reasonable success has also been observed with internal coordinate force fields developed by Zhurkin et al. (1980, 1991). With internal coordinate molecular mechanical force fields, solvent is rarely included explicitly and is instead modeled via a simple implicit solvent model. Effective dielectric treatments such as the distance-dependent dielectric ($\epsilon = 4r_{ij}$) or sigmoidal dielectric functions are commonly applied (as discussed in more detail in UNIT 7.9). More recently, increases in computer power have led to the incorporation of more accurate and implicit solvent models, such as finite difference Poisson-Boltzmann methods (Zacharias and Sklenar, 1997). Although it is not directly possible to mix and match the internal coordinate force fields with those designed for all-atom simulation, the FLEX force field has been shown to agree well with the all-atom nucleic acid force field described by Cornell et al. (1995) and Flatters et al. (1997). For more information on the differences between internal coordinate and all-atom representations, see UNIT 7.8.

For simulations where each atom is free to move (known as all-atom simulation), force fields have steadily and continually improved in recent years. For simulations including explicit solvent and a proper treatment of the electrostatic interactions, the most widely applied force fields for nucleic acids have been the Cornell et al. (1995) force field, the MacKerell force fields in CHARMM (Mackerell et al., 1995), and the force field in GROMOS (van Gunsteren and Berendsen, 1987). The latter force field performs poorly in nanosecond-length simulation (Tapia and Velazquez, 1997), presumably due to improper treatment of the long-range electrostatic interactions. The tendency to fail when the long-range electrostatic interactions are not properly included or omitted appears to be a general property of most of the available empirical force fields (Cheatham et al., 1995). As discussed in UNIT 7.9, to treat the electrostatic interactions correctly, it is recommended that a smooth cut-off method be applied (such as an atom-based force-shifted cut-off) or that the electrostatic interactions be fully included via an Ewald treatment.

Each of the currently available force fields for nucleic acids has various strengths and weaknesses. Direct comparison of all is difficult since, to date, there has not been a published systematic study of each, with consistent simulation protocols, to benchmark their relative performances. Despite this caveat, based on the authors' analysis of the literature and the authors' experience with many of these force fields, some insight can be found. The earlier MacKerell force field (Mackerell et al., 1995), also referred to as the all22 force field in CHARMM, accurately represents canonical A-form DNA structures. This force field does not properly stabilize canonical B-form DNA structures since slow transitions to A-DNA are seen in solutions with low salt conditions (Norberg and Nilsson, 1996a; Feig and Pettitt, 1997; MacKerell, 1997, 1998). This has been remedied in a more recent parameterization (all27; Foloppe and MacKerell, 2000; MacKerell and Banavali, 2000) that does a better job on B-DNA structures, albeit with less sequence-specific minor groove narrowing than expected. More consistent behavior with nucleic acids is observed with the Cornell et al. force field in conjunction with a particle mesh Ewald treatment of the electrostatic interactions. With this force field, spontaneous A-DNA to B-DNA transitions are seen with a variety of sequences as expected (Cheatham and Kollman, 1996, 1998), and B-DNA to A-DNA transitions are observed with phosphoramidate-modified backbones (Cieplak et al., 1997), consistent with those seen in experiments. This observation is very exciting since it suggests that conformational sampling is not overly inhibited under truly periodic boundary conditions for DNA in explicit solvents. The force field also predicts A-RNA to be stable, although B-RNA to A-RNA transitions do not occur spontaneously (Cheatham and Kollman, 1997b). This brings up the issue of poor sampling of RNA in nanosecond-length simulation since B-RNA is stable for >10 nsec unless concerted changes in the sugar pucker are forced. This is likely due to the larger barriers to conformational transition due to 2'-O interactions with the backbone and larger barriers to sugar repuckering. With this force field (and likely others), modelers should be aware that sampling is more limited in molecular dynamics simulations of RNA. The Cornell et al. force field also reproduces sequence-specific structures, such as the expected TpG step bends in the major groove or the narrowing of the minor groove in polyadenine (A-tract) regions, and shows good agreement with crystal data (Young et al., 1997b). Expected structural differences between DNA-DNA, DNA-RNA, and RNA-RNA duplexes are well modeled, as are modified nucleic acids such as phosphoramidates (Cieplak et al., 1997) and photo-damaged DNA (Miaskiewicz et al., 1996; Spector et al., 1997).

More recently, the Cornell et al. force field has also demonstrated the ability to model (with the same force field and simulation protocol) changes in nucleic acid structure that result from changes in the solvent environment. This includes the stabilization of A-DNA in a water and ethanol solution (Cheatham et al., 1997) and spontaneous B-DNA to A-DNA transitions in the presence of hexaammine cobalt(III) (Cheatham and Kollman, 1997a). Limitations of this force field include lower-than-expected sugar pucker, χ angles, and helical twist. This has been improved in a more recent parameterization (called parm98 or parm99) of the dihedral terms (Cheatham et al., 1999). Another force field worthy of notice is the BMS (Bristol Meyer Squibb) force field for nucleic acids (Langley, 1998), which was explicitly parameterized in order to properly simulate A-DNA/B-DNA equilibria under various conditions including water/ethanol and high salt.

Note that with all of these force fields the terminal phosphates are generally not included and, moreover, there is a distinction between internal, free, and 3'- or 5'-terminal residues. With each force field or simulation program, there are procedures to handle this distinction either through the use of different names for the residues (as in AMBER) or using various procedures designed to patch the terminal residues to shift charge and delete the terminal phosphate group (as in CHARMM).

Given the constantly changing landscape, the limited application of some of these force fields, and variations in the applied methods, it is difficult to evaluate which force field is “best”. Experience plays a part in the selection of a force field and, therefore, careful evaluation of the published reports is important. A final point is that not all force fields are compatible with a given simulation code. The Cornell et al. force field is released with AMBER, the MacKerell force fields with CHARMM, and other force fields with other codes such as GROMOS or Insight/Discover. Since CHARMM and AMBER use a similar molecular mechanics potential and equivalent Lennard-Jones combining rules, these force fields can be interconverted; this is not easily possible for GROMOS, which uses geometric-mean combining rules and has a different form for the bond- and angle-stretching terms.

Besides choosing which force field to use, an additional question is what to do if parameters are missing. This is a difficult question to answer in general terms, however, as most force fields have a specific protocol that should be followed to develop new parameters. For example, with the Cornell et al. force field, new intramolecular parameters are chosen by analogy to be consistent with existing parameters, general van der Waals parameters are obtained from simulations of neat liquids (e.g., OPLS; Jorgensen et al., 1996), and restrained electrostatic potential (RESP) fit charges from *ab initio* calculations are obtained for each new nucleotide, residue, or substructure. For more information about adding missing parameters, one should read all of the relevant force field literature and search for guides or repositories of existing parameters on the Internet. It is also possible to pose questions to the various e-mail reflectors for each program to ask if other investigators have already developed parameters for the system of interest. Additionally, it may be appropriate to contact the corresponding author of the force field papers for more information on the specific protocol for developing new parameters.

Balance is also an important requirement for a given force field; this means that in addition to accurately modeling the intra-DNA interaction, it is important to have a balanced representation of the DNA with solvent. The current force fields (Cornell et al., MacKerell, and Langley) appear reasonably balanced. In addition to balance of the DNA with the explicit solvent, these force fields, in general, allow simulation of protein systems. Reasonable representation of protein–nucleic acid structures has been seen with all three of the major force field derivatives (Cornell et al., MacKerell, and BMS; Wang et al., 2001). In addition to being balanced, the force fields do not appear to be overly sensitive to small changes in the force field in nanosecond-length simulation (Cheatham and Kollman, 1998). For example, with addition or removal of water, use of modified water that diffuses two times as fast as TIP3P, various low-salt environments, or other small changes, there does not appear to be a major systematic alteration in the observed structure. However, there are clear issues with the imposition of periodicity, as discussed in *UNIT 7.9*.

Setting Up a Nucleic Acid System With Explicit Water And Counter-Ions For Molecular Dynamics or Monte Carlo Simulation

A molecular dynamics (MD) or Monte Carlo (MC) simulation (see *UNIT 7.8*) is generally broken up into two sequential phases. The initial part of the simulation is the equilibration phase, and this is followed by the production or sampling phase. The equilibration phase includes the beginning parts of the simulation, often including both minimization and dynamics, performed in order to obtain a “stable” simulation, as measured by a varied series of parameters such as the root-mean-squared atomic deviation from the starting structure, the temperature, or the total molecular mechanical energy. Once the simulation is stable, the production phase is performed. The production phase is the part of the dynamics or sampling that is extensively analyzed. The omission of the equilibration

phase from this detailed analysis is necessary to avoid bias. The precise definition of what the production and equilibration phases entail are somewhat ambiguous. For example, when one parameter, such as root-mean-squared deviation (RMSd) of the solute to its starting structure or the temperature, has stabilized around an equilibrium value, this does not imply that all variables have equilibrated. Whether a given observable property has fully equilibrated depends on how long that observable takes to relax from its initial value to its equilibrium value. Many observable properties, such as the distribution of ions or sampling of thermally accessible conformational substates, take considerable amounts of time to equilibrate, times often longer than those that current state-of-the-art simulations can achieve. Therefore, when discussing equilibration, it should be referenced to a particular observable property (e.g., temperature, pressure, volume, potential energy).

Equilibration is necessary in molecular dynamics simulations to relax structural distortions and remove large forces that may bias the dynamics. This means that equilibration is necessary to thermalize the system to put a comparable amount of kinetic energy into each degree of freedom. When this is not done, large forces may result at the distortions that, in turn, lead to large collisions on the local scale and create local “hot spots”. These hot spots may move the structure in unrealistic ways. Therefore, the goal of the equilibration procedure is to relax the system as much as possible to avoid biasing it away from the starting geometry. This is generally done through a series of minimization and molecular dynamics simulations where the temperature or kinetic energy is gradually increased.

Setting up an initial in vacuo model, with or without inclusion of an implicit solvent model, is straightforward. Models can be generated in a variety of ways, usually based on known experimental structure. Issues related to model building are discussed in greater detail in *UNIT 7.5*. In molecular dynamics simulation of a model in vacuo, limited equilibration is necessary. As long as the structure is not significantly distorted, small distortions in the structure can be relaxed by short minimizations (~100 to 1000 steps). Generally, a simple first-order method, such as steepest descent, is applied first to remove the largest forces, followed by a faster directed minimization method, such as conjugate gradient minimization. Minimization is performed until the change in energy or gradient between each minimization step is small (~0.1 to 0.0001 kcal/mol). Careful thermalization of the system, via a series of short molecular dynamics simulations where the temperature is gradually raised, is usually not necessary for in vacuo simulations since the system equilibrates rapidly (due to significantly fewer degrees of freedom than corresponding simulations with explicit solvent). If explicit ions are included in the in vacuo simulation, more careful equilibration may be necessary to relax the ion atmosphere.

Adding explicit solvent significantly increases the complexity and computational cost, and also necessitates a more stringent equilibration protocol (see Basic Protocol). To perform a simulation with explicit solvent using an appropriate nucleic acid model starting structure (in vacuo), an initial configuration of the solvated model is necessary. Moreover, it is typically desirable to include at least enough explicit salt to neutralize the system and any excess salt as desired. Adding solvent is typically performed by first completely surrounding the desired system by a set of pre-equilibrated solvent “boxes” representing the coordinates of a unit cell of “bulk” water, and then deleting those waters that overlap the model or extend beyond the boundary of the system under consideration. The only problem with this approach is knowing how to place solvent inside interior cavities or at interfaces, such as a protein-DNA interface. Unfortunately, there is currently no clear consensus on how best to do this in the absence of experimental structural data. However, in the authors’ experience, the water structure relaxes rapidly and is able to diffuse efficiently even into tight interfaces where proteins bind DNA within short (100-psec)

simulations. In the absence of a pre-equilibrated solvent box, it is possible to simply add a crystalline representation of the solvent equally spaced at approximately the correct density; in this case, longer equilibration of the solvent may be necessary to fully relax the system and remove the crystalline bias.

As discussed in *UNIT 7.9*, enough water should ideally be added to not only fully hydrate the model but to represent bulk water some distance away from the model. However, adding water tremendously increases the cost of the simulation. Typically, when periodic boundary simulations are applied, ~ 8 to 12 \AA of water surrounding the model in each direction is appropriate, although more adventurous souls could use more or less water. Given the issues and potential for artifacts with nonperiodic systems, particularly for highly charged systems such as nucleic acids, nonperiodic boundary conditions are not recommended for nucleic acid simulations except under specific conditions. These include the desire to represent a minimally hydrated nucleic acid in vacuo or a very large nucleic acid where only a core part of the structure (surrounded by a blob of explicit water with stochastic boundary conditions) is of interest, as applied in various protein simulations (Brooks et al., 1985; Steinbach and Brooks, 1993) and has been applied in some simulations of nucleic acids (Norberg and Nilsson, 1996b; Mazur, 1998). Nonperiodic boundary conditions may also be appropriate for very large and irregularly shaped models where the regular shapes of periodic boundary conditions may require too much water or could, in principle, inhibit motion. However, in these cases, implicit solvent models are more likely appropriate. In the opinion of the authors, it is wise to avoid methods that include explicit solvent but also dampen the electrostatic interactions through the application of a distance dielectric function, since these methods tend to significantly dampen the conformational fluctuations.

For standard nucleic acid models representing a folded structure or 10 to 25 bp of a linear duplex, periodic boundary conditions are more appropriate. In principle, there is no reason to limit the shape of the periodic unit cell to a cubic shape, as any uniform space-filling shape is appropriate. Possible unit cell types are shown in Table 7.10.1. For long linear duplexes, the most appropriate unit cell might be the orthorhombic or hexagonal unit cells with one of the dimensions longer than the other two. This allows reasonable solvation of the duplex without large uninteresting regions that contain only solvent. However, an important consideration is that the rotational correlation time of small duplexes in solution is in the nanosecond time range. Therefore, during the dynamics in this type of orthorhombic box, the model may rotate to span the short edge of the box where the model can then interact directly with its periodic image (assuming a truly periodic method). This can also happen with constant pressure simulations where each box length is free to change. As shown in Figure 7.10.1, this can, in principle, lead to distortion of the model structure (as observed in long simulations of B-RNA in solution; Cheatham, unpub. observ.). This rotation can, in principle, be removed (at the expense of adding an uncorrected net torque to the system, which can lead to artifacts) or be inhibited by restraining the top and bottom of the duplex with weak restraints (which may inhibit bending). An additional issue is that interactions with periodic images in these long narrow boxes could, in principle, inhibit bending. In practice, inhibition of bending does not seem to be a major issue. Young and Beveridge (1998) reproduced expected sequence- and salt-specific bending in various phased A-tract models based on simulations of 24-mers in long rectangular boxes. To avoid the possible bias from inhibited rotation, more voluminous cubic boxes can be applied. This, however, leads to much more water in the corners than is necessary. Therefore, more modelers have shifted towards using more “spherical” unit cells, such as the 14-sided truncated octahedron (Allen and Tildesley, 1987) and 12-sided rhombic dodecahedron (see Figure 7.10.2). These unit cells limit the volume while maintaining distance between periodic images. Of course, adding the

Table 7.10.1 Standard Unit Cells Appropriate for Periodic Boundary Conditions^a

Restrictions on unit cell parameters, Volume

Cubic, $a = b = c$, $\alpha = \beta = \gamma = 90.0^\circ$, $V = a^3$

Tetragonal, $a = b$, $\alpha = \beta = \gamma = 90.0^\circ$, $V = ca^2$

Orthorhombic, $\alpha = \beta = \gamma = 90.0^\circ$, $V = abc$

Monoclinic, $\alpha = \gamma = 90^\circ$, $V = abc \times \sin(\beta)$

Triclinic, No restrictions, see legend

Hexagonal, $a = b$

Rhombohedral (trigonal), $a = b = c$, $\alpha = \beta = \gamma < 120.0^\circ$, $V = a^3 \times [1 - \cos(\alpha)] \times [1 + 2\cos(\alpha)]^{1/2}$

Octahedral (truncated octahedral), $a = b = c$, $\alpha = \beta = \gamma = 109.47122063449^\circ$, $V = (4(3)^{1/2}/9)a^3$

Rhombic dodecahedral, $a = b = c$, $\alpha = \gamma = 60^\circ$, $\beta = 90.0^\circ$, $V = (1/2)^{1/2}a^3$

^aRestrictions on the unit cells lengths (a, b, c) and angles (α, β, γ) are presented along with the volumes for a variety of simulation cells. The volume of a triclinic cell is $V = abc \times [1 - \cos(\alpha)^2 - \cos(\beta)^2 - \cos(\gamma)^2 + 2\cos(\alpha)\cos(\beta)\cos(\gamma)]^{1/2}$.

solvent is a little more tricky in nonorthorhombic unit cells since, when overlaying a larger solvent box, it is not a trivial procedure to remove waters outside the cell by simply checking if the water has coordinates larger than the box in a given dimension. In practice, a simple solution is to keep a set of the original coordinates with solvent, perform periodic imaging with the new unit cell type, and then, by comparison with the saved coordinates, delete any waters that have moved.

Although, in principle, adding salt is as easy as adding solvent, it is slightly more complicated in practice. In an ideal case, random waters might be replaced by explicit salt ions up to the desired concentration and then molecular dynamics or Monte Carlo simulation can be performed to equilibrate the salt. However, unlike water, which is at a relatively high concentration and equilibrates rather rapidly, specific association of ions to the nucleic acid and relaxation of the ion distribution, in principle, may take a significant amount of time. Therefore, this approach may be impractical. This is particularly true for

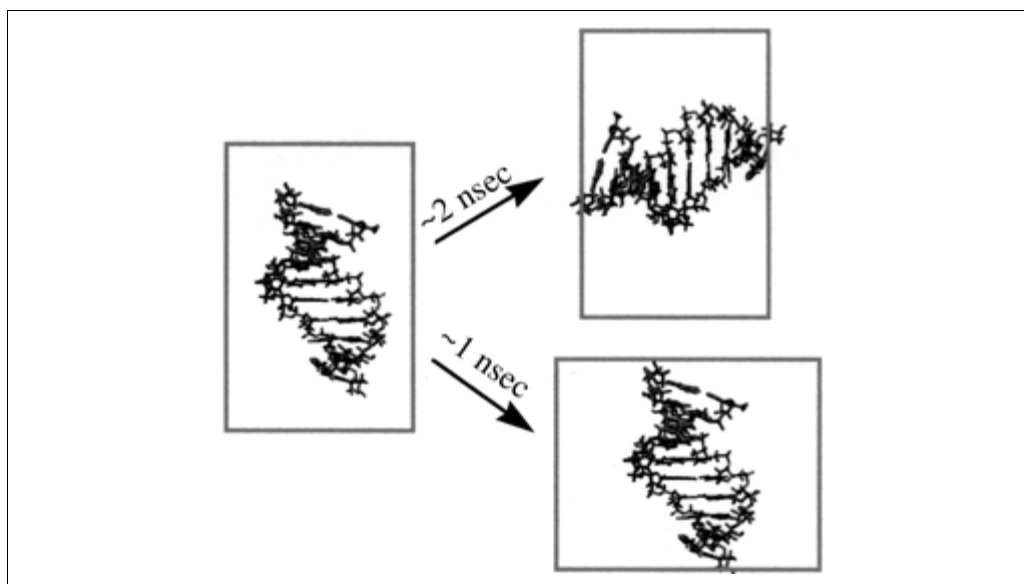


Figure 7.10.1 Orthorhombic unit cells and duplex rotation or unit cell size changes.

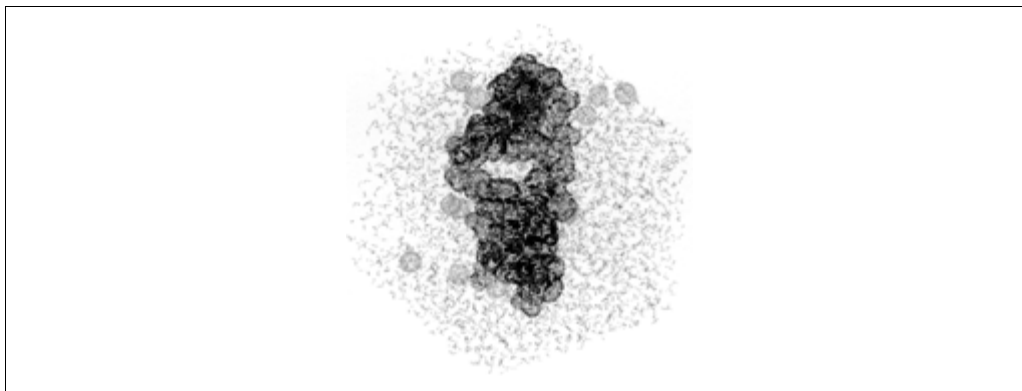


Figure 7.10.2 A model B-DNA/Z-DNA junction in a solvated rhombic dodecahedron unit cell after ~2 nsec of molecular dynamics with particle mesh Ewald (PME) in CHARMM.

the closely associated net-neutralizing counter-ions and multivalent ions that tend to have long water-exchange lifetimes and less rapid diffusion. Given the less rapid equilibration of multivalent ions, it is wise to make sure that each of these ions is sufficiently hydrated. This is wise because any direct binding of a multivalent ion may not exchange with water during nanosecond-length simulations. In the absence of any explicit information suggesting direct binding of an ion to the nucleic acid, placing ions that are directly bound (without bridging water) should be avoided, otherwise, the structure may distort under the influence of the bound ion. Placing ions within a hydration shell is reasonable since, in many cases, the interaction of a cation with a nucleic acid involves bridging water, such as with magnesium (Buckin et al., 1994) or barium (Sternglanz et al., 1976). In spite of this, specific interaction of ions with nucleic acids are seen. For example, ions have been observed in the minor groove of A-tract B-DNA (Hud and Feigon, 1997; Shui et al., 1998), interacting with the bases in the major groove of A-form structures (Robinson and Wang, 1996), or involved in phosphate interactions. Direct interaction is also seen with divalent ions in RNA that are known to stabilize the tertiary structure.

Before solvating the system, it is advisable to place the net-neutralizing counter-ions. If available, structural information regarding the placement of ions can be used as an initial guide. However, more often than not, this information is not available. The net-neutralizing counter-ions are added to balance the charge on the phosphates. Sodium ions are typically the ion of choice because of their fairly rapid diffusion, small size and atomic number, and relatively rapid exchange times. Precise placement is not a major issue. Reasonable molecular mechanical potentials exist for treating these ions, and a variety of parameterizations are in common use (Straatsma and Berendsen, 1988; Aqvist, 1990; Smith and Dang, 1994). In the absence of structural information, a simple procedure commonly used is to place the ions some distance ($\sim 5 \text{ \AA}$) from the phosphate along the bisector of the phosphate oxygens. This works reasonably well for B-DNA structures since the phosphates are regularly spaced and the vector from the phosphate to the bisector points away from the surface of duplex DNA. For A-form structures or folded structures, the phosphates may be closer together and the vector from the phosphate to the bisector may point to interior regions or other phosphates, which can lead to overlapping ions. These initial guesses can be easily relaxed for more favorable positions that avoid the overlap by using a quick in vacuo minimization with the nucleic acid held fixed. However, during the in vacuo minimization, it is important to avoid the direct approach of the ions to the nucleic acid, which will happen because water has not been included at this stage and the ion parameters for use in explicit water have a small van der Waals radius and large charge. This can be accomplished by altering the ion parameters to represent the ion as being effectively hydrated (Singh et al., 1985). This can be done by increasing the

van der Waals radius ($r_{\min} \approx 5 \text{ \AA}$) and decreasing the well depth ($\epsilon \approx 0.1 \text{ kcal/mol}$). For simulation with sodium ions (Na^+), given their reasonable diffusion, this procedure leads to reasonable interaction with the nucleic acid. With folded structures, care should be taken to make sure an ion is not buried where it should not be.

A more elaborate and perhaps better means to place the ions might be to use the electrostatic potential as a guide, such as that calculated from Poisson-Boltzmann treatments. For example, the program GRASP (<http://www.honiglab.cpmc.columbia.edu/grasp>) can display the electrostatic potential and ions can be placed in favorable positions. Alternatively, the individual grid elements can be investigated and ions placed based on grid elements with low energies. A simple method along these lines is to place the ions, based just on the electrostatic potential, on a grid representing the Coulombic potential of the nucleic acid interacting with an ion charge (q_{ion}) at each grid point (\vec{r})

$$E_{\text{grid}} = \sum_j^{\text{atoms}} \frac{q_j q_{\text{ion}}}{\epsilon |r_j - \vec{r}|}$$

where E_{grid} is the interaction energy of the ion with the grid element, q_j is the charge on each atom, r_j is the position of each atom, and ϵ is the dielectric constant. This is typically done assuming uniform screening ($\epsilon = 1$), although an effective dielectric could also be used. For placing ions around a specific group, such as a phosphate, a small grid is typically created. To avoid van der Waals overlap, the van der Waals energies can also be calculated for each grid element. The ion is then placed at the grid element with the lowest energy. To place all the ions, a larger and coarser grid can be built over the whole model (deleting grid elements within the solvent-accessible surface area or including van der Waals energy), and then the ions can be iteratively placed at low-energy grid points based on the energy, reevaluating the energy at each grid element after the placement of each ion. Alternatively, for placing positive ions in a previously solvated simulation, the DNA to water oxygen interaction energy can be used as a guide, with low interaction energies representing favorable positions in which the modeler can swap ions. The standard MD codes include procedures that allow placement of ions based on the phosphate bisector or other empirical rules, and some of these programs also have facilities for placing ions based on an energy grid. After addition of these ions, the system can be explicitly solvated. After this is done, it is a good idea to check the model structure to make sure no buried water appears where it is not appropriate and that the ions are explicitly and completely hydrated (unless directly bridging).

After placing the net-neutralizing ions, any excess salt can be placed after solvating the system by replacing random waters some distance away from the nucleic acid and other ions. The key question is how much excess salt should be added. Adding excess cation also means that balancing anions should be added. Since the system is being modeled at an atomic level with unit cell lengths of ~ 25 to 100 \AA , very small changes in the unit cell size or the number of counter-ions have a large influence on the effective molarity. For a given ion, the molarity as reported in simulation literature is based on either the total amount of a specific ion present in the simulation (not just excess salt) or the total number of excess ions per the total volume (converted to moles/liter). Since the system is not a bulk macroscopic system, this concentration may be much higher than expected under periodic boundary conditions. One can also refer to the ionic strength of the system; this includes all the ionized groups (including phosphates). Given the small size of the unit cells and relatively large number of phosphates to added salt, these simulations are most often performed at high ionic strength. Despite the sensitivity of molarity to unit cell size, when looking at monovalent ions, there is very little salt dependence on dynamics or structure over the range of no salt (including no net-neutralizing salt) to $\sim 1 \text{ M}$ salt in

1-nsec-length simulations with the Cornell et al. force field (Cheatham and Kollman, 1998). It is not until high salt concentrations (>3 to 4 M) are reached that transitions in DNA duplex structure are seen; these transitions have been observed in simulations with the BMS nucleic acid force field (Langley, 1998). Divalent and multivalent ions, on the other hand, have much more direct influence on structure. For example, only four $\text{Co}(\text{NH}_3)_6^{3+}$ ions are necessary to observe B-DNA to A-DNA transitions with the Cornell et al. force field (Cheatham and Kollman, 1997a). Magnesium may also effect bending (Young and Beveridge, 1998).

Equilibrating Simulations With Explicit Solvent

After generating initial ion and solvent positions, it is necessary to equilibrate the system. This relaxes the system to the expected density and allows the water and ions to react to the presence of the nucleic acid. Minimization to remove unrealistic energies is an essential first step to this process; however, it is not sufficient. Given the multiple minima problem (UNIT 7.8) and, moreover, the fact that minimized water or “ice” is not what is really desired, it is necessary to sample possible configurations via molecular dynamics or Monte Carlo simulation. Since the solvent was most likely placed suboptimally, there may be holes or gaps in the solvation. Furthermore, the “pre-equilibrated” water will not have reacted to the presence of the nucleic acid. Thus, the system will most likely not be at the correct density. To remedy this, constant pressure equilibration under periodic boundary conditions is likely necessary. The initial solvent (and ion) equilibration is the most important part of any equilibration protocol prior to production MD. Given appropriate simulation methodologies, if the solvent and ionic atmosphere is well equilibrated, the simulation will likely be stable. In this case, the precise and intricate details of the remainder of the equilibration protocol are likely to be unimportant. This has been shown in molecular dynamics simulations of a DNA duplex where, after equilibration, there was little observable effect of varied ion placement when comparing three different mechanisms for placing sodium counter-ions (Young et al., 1997a). It should be noted that “equilibration” in this context refers to the generation of a more reasonable solvent structure and initial configuration that (1) does not contain local hot spots with unreasonable forces, (2) is at the correct pressure and density, and (3) has a reasonably stable potential energy. This equilibration does not refer to complete equilibration of the nucleic acid model, a process that may take significantly longer.

A standard procedure (see Basic Protocol) is to first perform minimization to remove any large energies (which will lead to initially large forces), and then perform ~25 to 100 psec of dynamics of constant pressure, with the nucleic acid held fixed or restrained to the initial model structure to relax the water and ion environment. The progress of the equilibration procedure is typically monitored by plotting the potential energy, density, and pressure. Equilibration is thought to be complete when these (and other) values have stabilized (see Figure 7.10.3). After this phase, minimization is performed on the entire system with the restraints on the initial nucleic acid model structure gradually reduced. Then, dynamics are performed on the entire system, slowly raising the temperature.

Various modelers use different protocols (for examples, see Cheatham and Kollman, 1997b; MacKerell, 1997; Norberto de Souza and Ornstein, 1997; Young et al., 1997b), although, in practice, all these protocols seem to work consistently well. For more complicated systems, such as those involving high concentrations of salt or mixed solvent (such as ethanol and water), longer equilibration protocols are necessary. Note that this type of equilibration protocol tends to support water and ion conformations that stabilize the initial model structure. This can inhibit conformational transitions to other structures, such as B-DNA to A-DNA transitions in high salt, since the initial configurations are optimized to the “fixed” initial structure (Langley, 1998). A final note is that constant

pressure simulation methodologies are plagued with pitfalls. If the forces restraining or fixing the initial conformation of the nucleic acid are not properly included with the calculated pressure, the pressure may be overestimated, leading to box expansion upon pressure scaling in longer equilibration simulations. For a more detailed discussion of this and other issues, see Cheatham and Brooks (1998).

EQUILIBRATION

This protocol describes the constraint/restraint of solute, relaxation of restraints, and equilibration (without restraints). Minimization is performed for ~50 to 10,000 steps, although less minimization (~200 to 500 steps) may be acceptable. Initial equilibration in molecular dynamics simulation takes ~10 to 100 psec. It is important to avoid large force constants when applying harmonic restraints in molecular dynamics, since these may lead to high frequencies and require shorter time steps for proper equilibration. Force constants in the range of 5.0 to 15.0 kcal/mol/Å² are reasonable.

This protocol is intended to serve only as a guide. The primary literature, program manuals, and available resources on the Internet should be consulted for more information. For AMBER, see <http://www.amber.ucsf.edu/amber/tutorial>.

Constrain or restrain solute (optional for in-vacuo simulation)

Minimization:

- 1a. If necessary, turn off SHAKE constraints (that fix bond lengths involving hydrogen).
The need for this step depends on the minimizer, force field, and SHAKE algorithm used.
- 2a. Perform steepest descent minimization until energy change at each step is less than ~1.0 kcal/mol.
- 3a. Continue with conjugate gradient until energy change is less than ~0.1 kcal/mol.
- 4a. If initial dynamics “blow up” (e.g., through SHAKE failures or sudden large increases in energy), perform additional minimization.

Molecular dynamics:

- 1b. Apply an integration time step of 1 to 2 fsec for most available molecular mechanics force fields.
- 2b. Apply SHAKE on hydrogen atoms if using a rigid three-point water model.
- 3b. Apply constant pressure if using periodic boundaries in explicit solvent (*UNIT 7.9*).
- 4b. Maintain desired kinetic energy or slowly raise (ramp up) the kinetic energy to desired values. Perform molecular dynamics simulation for ~10 to 100 psec.

Longer equilibration times are likely necessary with high salt conditions, multivalent ions, mixed solvents, or slowly diffusing solvents.

Relax restraints (optional)

5. Perform minimization only (~50 to 5000 steps) with restraint force constants gradually moved to zero, or perform cycles of minimization (~50 to 5000 steps) and dynamics (~1 to 50 psec) with restraint force constants gradually moved to zero (i.e., 15 to 10 to 5 to 0.0 kcal/mol/Å²).

Perform equilibration without restraints

6. Perform minimization or molecular dynamics as described above.

- Make certain that the overall rotational and translational kinetic energy is removed after initial velocity assignment and at periodic intervals as necessary (Harvey et al., 1998).
- If molecular dynamics blow up due to SHAKE failure or large energy change, try more minimization. If dynamics continue to fail, try decreasing the integration time step. If it continues to fail, look for a strong overlap of atoms and/or improper treatment of the electrostatic interactions.
- Monitor potential energy and the root-mean-squared deviation (RMSd) from starting structure, temperature, pressure, density, and volume in order to judge progress of equilibration. Begin production dynamics when these and other interesting observables appear to stabilize.

At this point it is possible to change from an NPT (constant pressure, constant temperature) ensemble to an NVE (constant volume, constant energy) ensemble.

ANALYZING THE RESULTS

After equilibration, production simulations are run for as long as computationally feasible or necessary. Current state-of-the-art simulations of small solvated biomolecules (representing on the order of ~10,000 to 25,000 atoms) are performed for on the order of 1 to 25 nsec. At regular intervals, the configuration of the system (including the values of various energy terms and the atomic coordinates) should be saved and recorded to file(s). There are a variety of means to analyze the results. Analysis is performed not only for the purpose of extracting useful information about the structure but also to check the simulations for any aberrant behavior. In general, in order to obtain meaningful statistics when monitoring a particular observable, the simulation should be run on a time scale that is at least an order of magnitude longer than the correlation time of that particular observable. Some properties relax very quickly, such as various equilibrium properties of

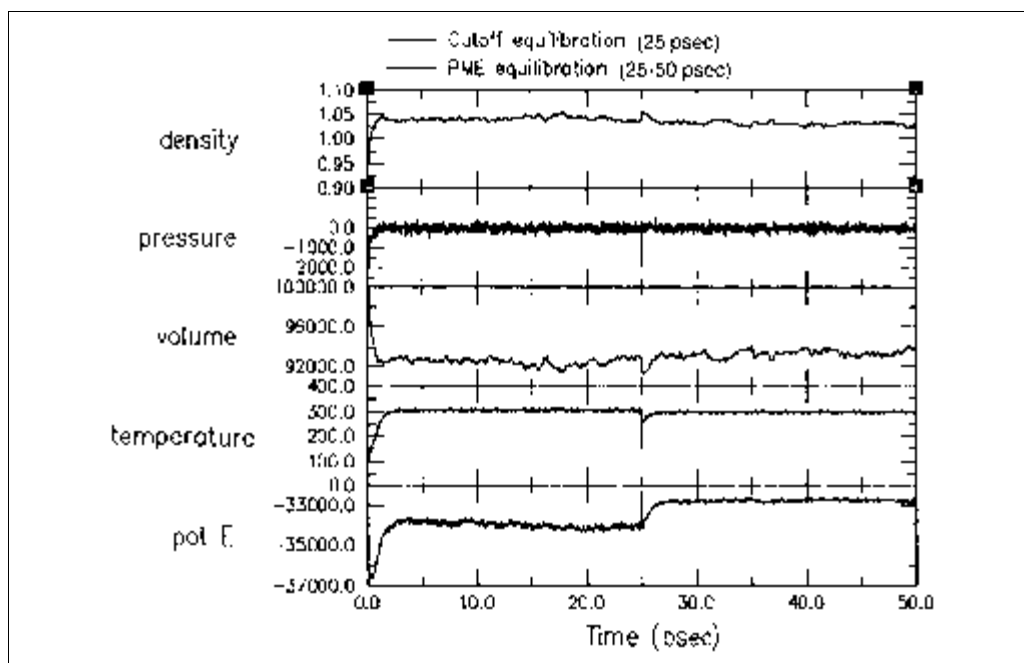


Figure 7.10.3 Graphs monitoring the equilibration of a simulation of a DNA (polyA-polyT 10-mer) in explicit water. The first part of the simulation (25 psec) was performed with a cutoff. The second part (25 to 50 psec) was performed applying the particle mesh Ewald method. The DNA was held fixed and only the ions and water were allowed to move.

water, i.e., density and average potential energy, which converge in short simulations (10 to 100 psec). Other properties, such as diffusion of salt, structural relaxation, or folding, may occur on a very long time scale. As discussed earlier, not all properties of a given system may fully equilibrate within the time scale of the simulation. Figure 7.10.3 shows a representative graph of various properties during a molecular dynamics simulation. From the graphs shown, it is clear that the simulation properties monitored fully converged during these short simulations.

Common properties to monitor include the RMSd from the starting and average structures (created by performing a straight coordinate average of RMSd fit configurations over a stable portion of the trajectory) and helicoidal parameters (*UNIT 7.5*), among others. In addition to investigating the time series, various correlation functions are also appropriate to extract information that can be compared more directly to experimental results. Although the production dynamics typically take the most time to perform, the major effort of any modeling project is often spent in analyzing the results. In large part, the type of analysis methods applied depends on what one is trying to learn from the simulation and what experimental data is available for comparison. There is not a specific protocol that can be summarized here. For more information and to get a better handle on various analysis tools, the primary literature should be consulted. A good source of information on modeling and analysis is provided by Leach (2001). Additional information is available with each of the simulation programs and on their respective Web sites. An excellent discussion of the major issues in simulation and analysis is provided by Mark and van Gunsteren (van Gunsteren, 1992; van Gunsteren and Mark, 1998). Although it is beyond the scope of this unit to discuss all the means for analyzing molecular dynamic trajectories in detail, an important tool worthy of discussion is the means to judge the importance of sampled conformations from a molecular dynamics or Monte Carlo simulation.

INEXPENSIVE METHODS TO ESTIMATE CRUDE RELATIVE FREE ENERGY DIFFERENCES

Given two different conformational states of the same molecule (using the same force field) sampled in Monte Carlo or molecular dynamics simulation, an estimate of the relative free energy can be obtained either by characterizing the set of configurations that represents each sampled state or by characterizing the minimum energy conformation that best resembles each sampled state (Kollman et al., 2000). This characterization involves estimating the relative free energy. In this context, the free energy is the sum of the enthalpy and a temperature-weighted entropy term. Determining the energy or enthalpy for a given state is relatively straightforward; it comes directly from the molecular mechanics energy function, either as an average over the configurations or as the minimum energy of a representative conformation for each state. As discussed in *UNIT 7.8*, it is not directly possible to compare molecular mechanical energies among different molecules (due to different zero-point energies) or with different force fields (due to possible different scales and different zero-point energies). Therefore, in this unit, the reference to relative energy and free energy differences are for the same molecule. For different molecules, other techniques may be more appropriate, such as free energy perturbation (discussed briefly at the end of this section). Typically, the solvent is not included explicitly (as discussed in more detail below) but can be represented implicitly.

In contrast to enthalpy, the entropy is less straightforward to estimate because it is an ensemble property. Although it can be calculated directly (at considerable cost), it is most often approximated and calculated independently for the solute and solvent. There are two basic methods for approximating the entropy; both comprise translational, rotational,

and vibrational components. The translational and rotational components are calculated for a rigid rotor approximation or by some other means. The vibrational component can be estimated via two methods. The first involves the use of a “representative” minimum energy conformation. For this conformation, the normal modes of vibration are calculated using a harmonic approximation. These normal mode frequencies can then be used to estimate the vibrational components of the entropy based on the local fluctuations in the neighborhood of the minimum energy conformation. This can give crude estimates, assuming that (1) the conformation is truly at the energy minimum, (2) the single minimum-energy conformation represents the state of interest, and (3) the anharmonic effects are small. However, the entropy typically involves more than local fluctuations within a given potential energy well for macromolecules, such as the entropy from larger scale conformational rearrangements. If the state of interest is characterized by a number of substates, which is likely the case (Poncin et al., 1992), the approximation of a single representative state may break down. Therefore, a set of representative states may be more reliable. However, counting the number of effective states and estimation of the energy based on a complete enumeration of the partition function necessitates reasonable sampling that may, in practice, not be feasible. The alternative procedure to calculate the vibrational entropy uses a quasiharmonic approach with vibrational frequencies estimated from the fluctuations observed during molecular dynamics. This allows estimation based only on the relatively important fluctuations in the representative set of states (Karplus and Kushick, 1981). Given the sampling difficulties, entropic effects are difficult to estimate and lead to the calculation of “crude” relative free energies. Note that these approaches are only valid (in practice) for estimation of the entropy of the solute.

In spite of the difficulties in estimating entropy, various groups have started to use data from a series of molecular dynamics simulations to estimate relative free energy differences. As a sample case study, consider the simulation of a canonical A-form RNA model and a canonical B-form RNA model, both of which are stable in multi-nanosecond-length state-of-the-art simulation including explicit solvent (Cheatham and Kollman, 1997b). Since both simulations are stable and no spontaneous conformational transition is seen, it is impossible to determine based on the MD results which conformation, B-RNA (a molecule that has never been experimentally observed) or A-RNA, is preferred. This is an important question since it would be valuable to better characterize the force field and to understand whether B-RNA is overstabilized by the Cornell et al. force field or is actually less favored than A-RNA. If A-RNA is indeed the energetically favored molecule, this implies that B-RNA to A-RNA transitions are not observed on a nanosecond time scale due to large conformational sampling barriers and insufficient sampling.

In general, the ability to rank the various “models” can be used to judge the utility or importance of a given model. Since the same sequence and force field is used, it is possible to directly compare the molecular mechanical energies (although, note that the solvent-solvent energies may have to be normalized if the two different simulations contain differing numbers of waters). The easiest way to estimate the relative free energy is to break up the total into contributions from the solvent (typically done implicitly) and the solute from the MD or MC simulation. Given a series of representative configurations from the dynamics, it is possible to determine the average intrasolute energy (or enthalpy). As mentioned, the entropy of the solute is a little more problematic to estimate, although the procedures specified above can be used. In general, a clear consensus on how best to estimate the entropic component has not emerged. One might consider making the assumption that the differences in entropy are largely represented in the solvation terms and not due to differences in the configurational entropy of the solute; however, this is likely invalid in most cases. With the A/B-RNA case study, in principle and as a first approximation, an estimation of the entropy can be obtained using the vibrational partition

function and a harmonic approximation to the normal modes to estimate the frequencies in representative minimum-energy conformations. A problem with this approach is that in vacuo minimization of nucleic acids (without the water) leads to distortion of the structure away from the structure represented in solution; therefore, the calculated entropy may not accurately represent the entropy as estimated from the various snapshots in the respective trajectories. It will, however, reasonably estimate the entropy of the gas phase model structure. For the model average structures of [CCAACGTTGG]₂ A-RNA and B-RNA (averaged over nanosecond portions of the respective trajectories taken at 1-psec intervals), minimization moved the models 1.9 Å and 2.4 Å, respectively, from the average structure. Estimation of the entropy difference using the normal modes favors A-RNA by ~3.0 kcal/mol (at 300 K); although this is not a large difference, it is significant (Cheatham, 1997).

The free energy of solvation can be estimated more directly. Under the assumption of linear response, a simple approximation to the solvation free energy in explicitly solvated simulations (also assuming that the bulk of this energy is represented by “close” waters, so that normalization for the total number of waters is not necessary) equates this free energy with half the solute-solvent interaction energy, $\frac{1}{2}E_{\text{solute-solvent}}$. A better estimate might be obtained by stripping the explicit water from each configuration and then performing a quick calculation on this conformation with an implicit solvent treatment. This will give an estimate of the solvation free energy. Recall that the implicit water models are typically parameterized to reproduce the free energy of solvation directly with a polarization component from Poisson-Boltzmann or a generalized Born treatment of the electrostatics and nonpolar contributions from a surface area term. This type of treatment has been applied to investigate a small turn-forming peptide based on long solvated MD trajectories (Bashford et al., 1997), and has recently been applied by various groups not only on the A/B-RNA case study but also the A/B-DNA equilibrium under various conditions, as well as a variety of other applications (Kollman et al., 2000). These techniques are very useful tools for post-processing MD or MC trajectories to give further insight.

For more accurate configurational free energy differences, it is necessary to explicitly sample the configuration of accessible conformations connecting the end points, or states of interest, in a single simulation or series of simulations. This is typically done by adding biasing potentials (umbrella sampling) to force sampling along a particular path (Valleau and Whittington, 1977). Typically, multiple simulations are applied with different biasing potentials along a given path or reaction coordinate, and the results from the various simulations are accumulated and unbiased through a procedure such as the weighted histogram method (Kumar et al., 1992, 1995; Roux, 1995). This procedure has been used to characterize protein folding (Boczek and Brooks, 1995), but has been used infrequently for nucleic acid simulation. The use of these biasing potentials requires some understanding of the reaction path between the two states of interest, and therefore is not straightforward and is very computationally demanding.

Calculating relative free energies of different molecules upon small chemical changes, free energy perturbation techniques, or thermodynamic integration techniques can be applied; for detailed reviews see Beveridge and DiCapua (1989) and Kollman (1993). Using data from an MD simulation can also give some qualitative insight (Radmer and Kollman, 1997).

SUMMARY

The methods and tools for accurate simulation of small nucleic acids in solution have advanced considerably in recent years. A summary of the highlights are presented in recent reviews (Cheatham and Kollman, 2000; Kollman et al., 2000). In general, when simulating a polyionic system such as nucleic acids, it is necessary to not only provide a proper representation of the long-range electrostatic interactions through atom-based force-shifted cutoffs or an Ewald treatment, but also include some representation of the surrounding environment (i.e., water and salt). Tremendous strides have been made in recent years, including accurate representation of A-tract bending, specific ion association, and sequence-specific structure and dynamics. The current generation of force fields still retains some systematic errors and clearly more computer power is necessary to begin to tackle larger-scale problems and longer simulation times. However, the future holds tremendous promise.

LITERATURE CITED

- Allen, M.P. and Tildesley, D.J. 1987. *Computer Simulation of Liquids*. Oxford University Press, Oxford.
- Aqvist, J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94:8021-8024.
- Bashford, D., Case, D.A., Choi, C., and Gippert, G.P. 1997. A computational study of the role of solvation effects in reverse turn formation in the tetrapeptides APGD and APGN. *J. Amer. Chem. Soc.* 119:4964-4971.
- Beveridge, D.L. and DiCapua, F.M. 1989. Free energy via molecular simulation. *Annu. Rev. Biophys. Biophys. Chem.* 18:431-492.
- Boczek, E.M. and Brooks, C.L. III. 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* 269:393-396.
- Brooks, C.L. III, Brunger, A., and Karplus, M. 1985. Active site dynamics in protein molecules: A stochastic boundary-molecular dynamics approach. *Biopolymers* 24:843-865.
- Buckin, V.A., Kankiya, B.I., Rentzeperis, D., and Marky, L.A. 1994. Mg²⁺ recognizes the sequence of DNA through its hydration shell. *J. Amer. Chem. Soc.* 116:9423-9429.
- Cheatham, T.E. III. 1997. Ph.D. dissertation. Realistic simulation of nucleic acids in solution. University of California at San Francisco.
- Cheatham, T.E. III and Brooks, B.R. 1998. Recent advances in molecular dynamics simulation towards the reliable representation of biomolecules in solution. *Theor. Chem. Acc.* 99:279-288.
- Cheatham, T.E. III and Kollman, P.A. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259:434-444.
- Cheatham, T.E. III and Kollman, P.A. 1997a. 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.
- Cheatham, T.E. III and Kollman, P.A. 1997b. Molecular dynamics simulations highlight the structural differences in DNA:DNA, RNA:RNA and DNA:RNA hybrid duplexes. *J. Amer. Chem. Soc.* 119:4805-4825.
- Cheatham, T.E. III and Kollman, P.A. 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, eds.) pp. 99-116. Adenine Press, Schenectady, N.Y.
- Cheatham, T.E. III and Kollman, P.A. 2000. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* 51:435-471.
- Cheatham, T.E. III, Miller, J.L., Fox, T., Darden, T.A., and Kollman, P.A. 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.
- Cheatham, T.E. III, Crowley, M.F., Fox, T., and Kollman, P.A. 1997. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci. U.S.A.* 94:9626-9630.
- Cheatham, T.E. III, Cieplak, P., and Kollman, P.A. 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.
- Cieplak, P., Cheatham, T.E. III, and Kollman, P.A. 1997. Molecular dynamics simulations find that 3' phosphoramidate modified DNA duplexes undergo a B to A transition and normal DNA duplexes an A to B transition. *J. Amer. Chem. Soc.* 119:6722-6730.

- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.* 117:5179-5197.
- Feig, M. and Pettitt, B.M. 1997. Experiment vs force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem.* 101:7361-7363.
- Flatters, D., Zakrzewska, K., and Lavery, R. 1997. Internal coordinate modeling of DNA: Force field comparisons. *J. Comp. Chem.* 18:1043-1055.
- Foloppe, N. and MacKerell, A.D. Jr. 2000. All-atom empirical force field for nucleic acids. (1) Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.* 21:86-104.
- Harvey, S.C., Tan, R.K.-Z., and Cheatham, T.E. III. 1998. The flying ice cube: Velocity rescaling in molecular dynamics simulations leads to violation of equipartition. *J. Comp. Chem.* 19:726-740.
- Hud, N.V. and Feigon, J. 1997. Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Amer. Chem. Soc.* 119:5756-5757.
- Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.* 118:11225-11236.
- Karplus, M. and Kushick, J.N. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 14:325-332.
- Kollman, P.A. 1993. Free energy calculations—Applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395-2417.
- Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., and Cheatham, T.E. III. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889-897.
- Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., and Rosenberg, J.M. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comp. Chem.* 13:1011-1021.
- Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A. 1995. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comp. Chem.* 16:1339-1350.
- 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.
- Lavery, R., Zakrzewska, K., and Sklenar, H. 1995. JUMNA (junction minimisation of nucleic acids). *Comp. Phys. Comm.* 91:135-158.
- Leach, A.R. 2001. *Molecular Modeling: Principles and Applications*, 2nd ed. Pearson Education Limited, Essex, England.
- MacKerell, A.D. Jr. 1997. Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties. *J. Phys. Chem.* B101:646-650.
- MacKerell, A.D. Jr. 1998. Observations on the A versus B equilibrium in molecular dynamics simulations of duplex DNA and RNA. *In Molecular Modeling of Nucleic Acids* (N.B. Leontis and J. Santa Lucia, eds.) pp. 304-311. American Chemical Society, Washington, D.C.
- MacKerell, A.D. Jr. and Banavali, N. 2000. All-atom empirical force field for nucleic acids. (2) Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* 21:105-120.
- Mackerell, A.D. Jr., Wiorkiewicz-Kuczera, J., and Karplus, M. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Amer. Chem. Soc.* 117:11946-11975.
- Mazur, A.K. 1998. Accurate DNA dynamics without accurate long-range interactions. *J. Amer. Chem. Soc.* 120:10928-10937.
- Miaskiewicz, K., Miller, J., Cooney, M., and Osman, R. 1996. Computational simulations of DNA distortions by a cis,syn-cyclobutane thymine dimer lesion. *J. Amer. Chem. Soc.* 118:9156-9163.
- Norberg, J. and Nilsson, L. 1996a. Constant pressure molecular dynamics simulations of the dodecamers: d(GCGCGCGCGCGC)₂ and r(GCGCGCGCGCGC)₂. *J. Chem. Phys.* 104:6052-6057.
- Norberg, J. and Nilsson, L. 1996b. Glass transition in DNA from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 93:10173-10176.
- Norberto de Souza, O. and Ornstein, R.L. 1997. Effect of warmup protocol and sampling time on convergence of molecular dynamics simulations of a DNA dodecamer using AMBER 4.1 and particle mesh Ewald method. *J. Biomol. Struct. Dyn.* 14:607-611.
- Poncin, M., Hartmann, B., and Lavery, R. 1992. Conformational sub-states in B-DNA. *J. Mol. Biol.* 226:775-794.

- Radmer, R. and Kollman, P.A. 1997. Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *J. Comp. Chem.* 18:902-919.
- Robinson, H. and Wang, A.H.-J. 1996. Neomycin, spermine and hexaamminecobalt(III) share common structural motifs in converting B- to A-DNA. *Nucl. Acids Res.* 24:676-682.
- Roux, B. 1995. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.* 91:275-282.
- Shui, X., McFail-Isom, L., Hu, G.G., and Williams, L.D. 1998. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* 37:8341-8355.
- Singh, U.C., Weiner, S.C., and Kollman, P.A. 1985. Molecular dynamics simulations of d(C-G-C-G-A)-d(T-C-G-C-G) with and without "hydrated" counterions. *Proc. Natl. Acad. Sci. U.S.A.* 82:755-759.
- Smith, D.E. and Dang, L.X. 1994. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.* 100:3757-3766.
- Spector, T.I., Cheatham, T.E. III, and Kollman, P.A. 1997. Unrestrained molecular dynamics of photodamage DNA in aqueous solution. *J. Amer. Chem. Soc.* 119:7095-7104.
- Steinbach, P.J. and Brooks, B.R. 1993. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.* 90:9135-9139.
- Sternglanz, H., Subramanian, E., Lacey, J.C.J., and Bugg, C.E. 1976. Interactions of hydrated metal ions with nucleotides: The crystal structure of barium adenosine 5'-monophosphate heptahydrate. *Biochemistry* 15:4797-4802.
- Straatsma, T.P. and Berendsen, H.J.C. 1988. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* 89:5876-5886.
- Tapia, O. and Velazquez, I. 1997. Molecular dynamics simulations of DNA with protein's consistent GROMOS force field and the role of counterions' symmetry. *J. Amer. Chem. Soc.* 119:5934-5938.
- Valleau, J.P. and Whittington, S.G. 1977. A guide to Monte Carlo for statistical mechanics. 1. Highways. In *Statistical Mechanics A: A Modern Theoretical Chemistry*, Vol. 5-6 (B.J. Berne, ed.). Plenum Press, New York.
- van Gunsteren, W.F. 1992. On the interpretation of biochemical data by molecular dynamics computer simulation. *Eur. J. Biochem.* 204:947-961.
- van Gunsteren, W.F. and Berendsen, H.J.C. 1987. Groningen molecular simulation (GROMOS) library manual. BIOMOS, Nijenborgh, Groningen, The Netherlands.
- van Gunsteren, W.F. and Mark, A.E. 1998. Validation of molecular dynamics simulation. *J. Chem. Phys.* 108:6109-6116.
- Wang, W., Donini, O., Reyes, C.M., and Kollman, P.A. 2001. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* 30:211-243.
- Young, M.A. and Beveridge, D.L. 1998. Molecular dynamics simulations of an oligonucleotide duplex with adenine-tracts phased by a full helix turn. *J. Mol. Biol.* 281:675-687.
- Young, M.A., Jayaram, B., and Beveridge, D.L. 1997a. 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.
- Young, M.A., Ravishanker, G., and Beveridge, D.L. 1997b. A 5-nanosecond molecular dynamics trajectory for B-DNA: Analysis of structure, motions and solvation. *Biophys. J.* 73:2313-2336.
- Zacharias, M. and Sklenar, H. 1997. Analysis of the stability of looped-out and stacked-in conformations of an adenine bulge in DNA using a continuum model for solvent and ions. *Biophys. J.* 73:2990-3003.
- Zhurkin, V.B., Poltev, V.I. and Florent'ev, V.L. 1980. Atom-atom potential functions for conformational calculations of nucleic acids. *Mol. Biol.* 14:1116-1130.
- Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A., and Jernigan, R.L. 1991. Static and statistical bending of DNA evaluated by Monte Carlo calculations. *Proc. Natl. Acad. Sci. U.S.A.* 88:7046-7050.

Contributed by Thomas E. Cheatham, III
University of Utah
Salt Lake City, Utah

Bernard R. Brooks
National Institutes of Health
Bethesda, Maryland

Peter A. Kollman
University of California
San Francisco, California