Published in final edited form as:

Biopolymers. 2013 December ; 99(12): . doi:10.1002/bip.22331.

Twenty-five years of nucleic acid simulations

Thomas E. Cheatham III¹ and David A. Case²

¹Department of Medicinal Chemistry, College of Pharmacy, University of Utah, 2000 East 30 South Skaggs 105, Salt Lake City, UT 84112

²BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers University, 174 Frelinghuysen Rd, Piscataway, NJ 08854

Abstract

We present a brief, and largely personal, history of computer simulations of DNA and RNA oligonucleotides, with an emphasis on duplex structures and the Amber force fields. Both explicit and implicit solvent models are described, and methods for estimating structures, thermodynamics and mechanical properties of duplexes are illustrated. This overview, covering about two decades of work, provides a perspective for a discussion of prospects and obstacles for future simulations of RNA and DNA.

1 Introduction

Computer simulations have had an important impact on our understanding of structures and dynamics of biomolecules. The “modern” era of nucleic acid simulations began about two decades ago, with the introduction of Ewald-based methods to handle electrostatic interactions in a consistent way in molecular dynamics simulations, and with the growth of numerical Poisson-Boltzmann calculations to provide insights at a more coarse-grained level where water and mobile ions are treated as a continuous distribution. Impressive advances since then in both numerical algorithms and in computer speed have greatly enhanced our ability to make critical comparisons to experiment on increasingly long time scales. Some of this history is recounted here. We write largely from a personal perspective, but with the goal of providing a framework for an evaluation of the prospects for, and obstacles to, making computer simulations increasingly accurate as a predictive and interpretative tool, and more generally useful in addressing biological problems.

2 Developments in nucleic acid force fields

Although there were many early attempts to describe nucleic acids using molecular mechanics force fields, it was not until the 1990’s that computer power was sufficient to run simulations long enough to really test the accuracy of the results. A number of published reviews outline the development of the “second generation” force fields for nucleic acids that occurred at this time, and many of these reviews also include critical assessment and validation of the force fields in simulations of nucleic acids.^{1–6} To enable reliable simulation of nucleic acids, key enablers were the development of particle mesh Ewald methods as an efficient means to treat long range electrostatic interactions,^{7,8} and the availability of efficient and parallelized versions of simulation codes.⁹

As the user community became more experienced with nucleic acid simulation, a number of deficiencies began to emerge. With the Amber nucleic acid force fields, the most notable deficiencies were a poor distribution of sugar pucker phases and a modest under-twisting of DNA duplexes. In an attempt to overcome these deficiencies, a large parameter scanning sweep on sugar and glycosidic torsion parameters was performed in a brute-force manner to

better understand how small changes the parameters altered the twist and sugar pucker distributions.¹⁰ This involved engineering small changes in various dihedral parameters which were then explored in ~5–10 ns molecular dynamics simulations to see their influence on the DNA structural properties and A to B transitions, with a focus on helical twist and sugar pucker distributions. This parameter scanning led to the AMBER *ff98* nucleic acid force field,¹⁰ which was tweaked a little bit further into the *ff99* force field.¹¹ Although both of these force field modifications improved the helical twist distributions slightly, the approach is not optimal. Small changes in one dihedral may influence other parameters and lead to structural artifacts, the changes were limited in scope, and the testing procedure only investigated the influence on the folded canonical DNA structural properties (and A–B transition rates) instead a broader structural exploration. Limitations in computer time prevented both more detailed structural investigation and testing, but also the application of more accurate high-level quantum mechanical (QM) calculations on larger and more relevant model systems. Although clearly not converged, and really only measuring sampling around the starting geometry (i.e. a small part of the accessible conformational space), the parameter scannings investigated were the longest simulations that were practical, and, the differences in the results were able to isolate changes to the dihedral potential that could improve the helical twist and sugar pucker distribution.

The next breakthrough's in the Amber nucleic acid force field development came from observations from relatively longer simulations on the 50–100 ns time scale in the early 2000's. For example, when investigating the influence of duplex length on d(GG)_n and d(CG)_n duplexes, and also in extended runs of DAPI bound to DNA¹² we (extending upon isolated observations of / transitions in the DNA backbone¹³) observed systematic overpopulation of = *trans* backbone geometries in simulations of nucleic acids. Rather than explore further dihedral parameter scanning, we took advantage of improvements in computational power and the efficiency of QM methods to perform full dihedral scans on more relevant model systems with better QM treatments and higher level basis sets. Towards this end, high level QM calculations were performed on models of sugars and phosphates, specifically a sugar-phosphate model¹⁴ and a sugar-phosphate-sugar model,¹⁵ which ultimately showed a misbalance in the molecular mechanical treatment that favored = *trans* too strongly and led to the *ff99-bsc0* parameterization.¹⁴ For simulation of canonical DNA and RNA structures, the *ff99-bsc0* parameterization has proven rather successful. For example, the Ascona B-DNA consortium (ABC) of different research groups applying MD simulation methods to DNA has investigated the sequence dependent structure and dynamics of tetrameric sequence repeats in DNA duplexes in 50–100 simulations now approaching over 1 μs for 18 different DNA duplex structures with the simulations providing new insight into genome stability and coarse-grained models of DNA.^{16–19}

For non-canonical structures, particularly those with loops or bulges, or flips, some anomalies have been noted. For example, with RNA, incorrect loop geometries, backbone substate populations and sugar pucker populations were observed in longer simulations. In addition to not being able to always maintain south puckers where found in RNA structures, multiple groups noticed a tendency for the RNA backbone to shift putting into the high-anti region which leads to an opening of the duplex structure into a ladder-like configuration. Again, QM methods at various levels were employed to improve the distribution using relevant model systems. The most tested modifications are the OL modifications referenced in Amber *ff11* and *ff12*.^{20,21} An alternative available with Amber is the Yildirim modifications (and also related modifications called TOR which alter / as well), although these have only been really validated and assessed in simulations of small oligonucleotides.^{22–24} To test out these and more recent force field modifications, a systematic assessment and validation of these newer modifications is underway on a large series of RNA tetraloop structures.

During these force field modification investigations a few lessons have been learned and there remain some issues to be resolved. Most notable is that small changes to a particular dihedral may lead to alteration in properties of related dihedrals, and may have unintended consequences. For example, the *ff99-bsc0* modifications tend to lock RNA sugar puckers mainly in the north, even with nucleotides in particular sequence contexts that prefer southern conformations. Moreover, the *bsc0* modifications tend to further destabilize *trans*. This suggests that to reliably improve the nucleic acid dihedrals, a more systematic approach across many dihedrals with simultaneous fitting may be more appropriate. Moreover, no longer do we fully support the idea that parameters are transferable between DNA and RNA, or between purines and pyrimidines. For example, the *ff99-OL* modifications (with or without *ff99-bsc0*) improve the modeling of RNA, but lead to issues with DNA, most notably with quadruplex structures. Therefore recent work has focused both on separate modifications for DNA and also improvements in the *ff99* backbone parameters.^{25,26}

In addition to issues with the nucleic acid force field, results are also dependent on the choice of solvent models and also ion parameters. The standard AMBER ions parameters released with the *ff94-ff99* force fields used AMBER “adapted for Lorentz-Bertholet combining rules” Aqvist monovalent cations with Cl⁻ anions from Smith and Dang, which, although adequate for NaCl simulation at low concentrations proved totally inappropriate for KCl which crystallized below physiological concentrations (~100 mM or above).²⁷ This led to the development of a newer set of monovalent ion parameters appropriate for particular water models.^{28,29} More recent work suggests potential misbalance between cations and phosphates that can be overcome by altering the van der Waals interactions for the specific interactions using the CHARMM NBFIX procedure to specifically define pair non-bonded interactions.³⁰ An additional ion parameterization of note is the refinement of Mg²⁺ ion parameters in order to better match coordination lifetimes; in our experience, despite the lack of explicit polarization, this simple Mg²⁺ model performs well in simulations of RNA.³¹ In addition the choice of water model could impact results, for example with SPC/E that tends to compact DNA duplexes to a greater extent than observed in simulations with TIP3P.³²

Key lessons learned are the need for long simulations, with or without the use of enhanced sampling (for example replica-exchange), on a wide diversity of model structures in order to assess and validate the simulation results. For example, although with the latest RNA force fields UUCG tetraloops may be stable on a 1–2 microsecond time scale, longer simulations on DE Shaw’s Anton machine at the Pittsburgh Supercomputing Center suggest irreversible unfolding and population of anomalous sub-states. Yet, in other cases, for example GNRA tetraloops, the simulations appear fairly robust in reproducing the experimental tetraloop structure.³³ Given the difficulties in completely sampling the conformational distribution of a RNA tetraloop, we have begun to study simpler systems where complete sampling is possible using standard or enhanced replica exchange MD simulations. Inspired by the work of Turner and co-workers, we have started to systematically evaluate the conformation distributions of RNA tetranucleotides, for example r(GACC) and r(CCCC).^{23,34} The MD sampling with the latest Amber RNA force fields (*ff12*) clearly show underpopulation of the expected A-like conformations and unexpectedly high population of anomalous conformations inconsistent with NMR data; this work highlights means to obtain well-converged conformational ensembles for these tetranucleotide model systems.³⁵ In addition to exploring a wide variety of conformational motifs, recent work has also begun to explore crystal simulations. The hope is that with further exploration of both small and larger nucleic acid models, spanning a large range of conformational motifs, we will be able to systematically explore nucleic acid structure and dynamics, expose force field deficiencies and continue to improve the additive nucleic acid force field models.

3 Implicit solvent models

The simulations described above use an explicit solvent model, in which all the atoms of the nucleic acid and its surrounding ion and water environment are explicitly represented. This can provide a faithful description of atomic reality, but a large fraction of the computing effort is devoted to the solvent. For example, in the “ABC” simulations described above,^{16–19} a typical DNA duplex (with 36 nucleotides) is comprised of about 1100 atoms, but the solvated system has about 36,000 atoms. Interest in faster simulations, and in models that might offer improved insight into solvent effects have driven interest in implicit solvent models, where only the atoms of the oligonucleotide solute are explicitly represented.

The simplest, and very powerful, implicit model is to treat water as a dielectric continuum. At low salt concentrations, the volume occupied by ions is small compared to that of the water molecules, and it is attractive to model the system with a dielectric near 80 everywhere, that is, to simply divide all electrostatic interactions by the dielectric constant. If the added ions are treated as hard spheres, one has the “primitive model” of electrolytes,³⁶ and there is a large literature on how such an ion-water environment would interact with polyions with the shape and charge distribution of RNA or DNA duplexes.³⁷

These primitive models have important limitations. First, they don’t (easily) provide a model for solvent forces, and do not, therefore, allow for simulations of conformational changes in the solute duplex. Second, the volume occupied by macromolecules is not generally small relative to the water volume, and it is not reasonable to assign a high dielectric constant to all space; rather, a division between high dielectric regions (in the solvent) and low dielectric regions (inside the macromolecular solute) is required, with a corresponding added complexity in handling electrostatic interactions in such a multi-dielectric environment. For biomolecular solutes of arbitrary shape, a numerical solution of the Poisson equation can be used to evaluate electrostatic interactions,^{38,39} The effects of mobile ions in solution can be approximately accounted for by assuming that the distribution of ions is proportional to $\exp(-ez / k_B T)$, where ϕ is the local electrostatic potential and z the ion charge.⁴⁰ This leads to the Poisson-Boltzmann (PB) equation, which has been widely employed in the exploration of electrostatic and salt effects in nucleic acids.^{40–43} It is a true implicit solvent model in the sense of providing the forces on the solute arising from the solvent polarization,⁴⁴ but until recently the computations have been too time-consuming to allow any routine use of this sort of model in molecular dynamics simulations. Rather, attention has been given to approximate models that incorporate many of the same physical ideas.

By far the most popular of these simpler models is the generalized Born approach,^{45,46} which builds on a simple picture of molecules as composed of atoms represented by overlapping spheres; the union of these spherical volumes represents the “interior” of the molecule characterized by a low-dielectric constant. Each atom is characterized by an effective radius, that describes its average distance from the molecular surface. The total solvation free energy can then be constructed from a sum over pairs of atoms. The entire procedure is analytical, so that molecular forces (and even second derivatives⁴⁷) can be obtained in an efficient calculation. There are a variety of arguments that can be used to justify the form of the approximations used in generalized Born theory, but it is perhaps best to think of it as an interpolation formula: at long distances, charge-charge interactions are screened by the solvent dielectric, whereas at short distances there is a penalty for burial of polar or charged groups away from solvent. A great deal of effort has gone into testing variants of these ideas against both numerical Poisson calculations and against experiment.^{48,49,50} Extensions have been made to include salt effects at the Debye-Huckel level of approximation.^{51,52}

The first extensive use of the generalized Born model in biomolecular MD simulations was to B-form DNA double helices^{53,54} (presented in part in the short-lived publishing experiment called *Biopolymers: Nucleic Acid Sciences*.) A systematic study of DNA duplex results using several GB models showed that most of the local sequence-dependent properties seen in explicit-solvent simulations are reproduced with GB models.⁵⁵ RNA structures have been more troublesome: A-form helices in generalized Born models (at least from Amber) tend to have major grooves that are too wide,⁵⁶ and even short single-stranded loops (such as stable UUCG or GNRA tetra-loops) are stabilized in conformations not seen experimentally, as discussed more fully above.

With generalized Born solvent models, it is feasible to study nucleic acid duplexes up to several hundred base-pairs. It is not obvious that the global properties of these longer sequences, such as their mechanical behavior, should match experiment all that well, since the parameterization of the force field and solvent model is based on the consideration of very small oligonucleotides. Figure 1 illustrates the results of one study of the mechanical properties of linear and circular DNA duplexes.⁵⁷ We computed normal mode of duplexes of various lengths, base sequences, and concentrations of added salt, exploiting the fact that the GB solvent model is fully analytic, so that constructing the Hessian matrix of second derivatives is a straightforward, if onerous, task.⁴⁷ The frequencies of the low-lying vibrational states, corresponding to the easiest distortions, can be compared to those from an elastic-rod model,^{58–63} and the elastic constants, such as bending, stretching and torsional rigidities, can be fit to the all-atom frequencies. The results are in good qualitative agreement with experiment, showing, for example that the bending rigidity decreases with increasing added salt, but that the stretching rigidity has the opposite behavior.⁵⁷ The mechanical constants of DNA mini-circles (of about 94 base-pairs, similar to that seen in DNA wrapped around nucleosomes) are very little different (for a given base sequence) than those for linear duplexes.

4 The MM-PBSA model

Although the implicit solvent models described above have many useful attributes, it generally seems to be the case (for both proteins and nucleic acids) that the ensemble of structures that is sampled is less faithful to experiment than are the ensembles derived from explicit solvent simulations. That realization has led to a strategy for energetic analysis that attempts to build on the strengths of both explicit and implicit solvent models to estimate free energy changes of conformational transitions. The basic approach is often called “snapshot analysis”: coordinates are saved at regular intervals from an explicit solvent simulation, but only the solute’s coordinates are used. Each solute snapshot is then immersed into an implicit solvent model for energy analysis. In this model, the contribution of snapshot i to the free energy of conformer x can be written as

$$G^x(i) = V_{ff}^x(i) + G_{solvation}^x(i) - TS^x(i) \quad (1)$$

Here $V_{ff}^x(i)$ is the value of the force field potential energy of snapshot i , sometimes referred to as the “gas-phase” potential energy, since no waters or other solvent molecules are present. (There is also a kinetic energy contribution to thermodynamic energies; in classical statistical mechanics this is just equal to $3k_B T/2$ for each atom, which will cancel when computing differences in free energies. Effects on the free energy due to changes in the pV terms are neglected.) The key feature that makes the MM-PBSA model “work” is that implicit solvent models provide an estimate of the free energy of transfer for gas to solvent, and thus no explicit estimates of the change in entropy of the solvent are required. The final, entropic, term in Eq. 1 concerns only the solute; this can still be a very difficult quantity to

model if the solute is large or floppy. Averages over snapshots (over i) then yield an estimate of the free energy of conformer x ; free energy *differences* are obtained by repeating the calculation for a different conformational state, *i.e.* by an “end-point” analysis. The virtue of this approach is that no pathway is needed for the conformational change; the main disadvantages are that implicit solvent models for $G_{solvation}^x(i)$ are of limited accuracy, and that it can be difficult to estimate the solute entropy terms.

Although the basic ideas here have many precedents, the first application of this model in its present form was to the A- and B-forms of RNA and DNA, where many details of the basic method are given.⁶⁴ There are also reviews summarizing many of the applications of this model.^{65–70} As an example, Fig. 2 shows some results from the original MM-PBSA paper,⁶⁴ comparing estimates of the free energy differences between A- and B-form helices for DNA and RNA. The top line gives the bare, (or gas-phase) electrostatic energy difference between A and B helices. Since the phosphate groups are somewhat closer together in an A-form helix, this value is large and negative for both DNA and RNA. Most of this bare interaction is screened by water, as represented by the positive “PB” number in the second row. The sum of the first two rows represents the effective electrostatic interactions, and it is clear that the implicit solvent theory used (which is Poisson-Boltzmann in this example) needs to be accurate to get a meaningful sum. The third row shows results for the van der Waals (or Lennard-Jones) contributions; the difference between DNA and RNA here primarily reflects that fact that there is not quite enough space in the B-form helix to accommodate the “extra” 2 OH of the ribose. The fourth row shows the bond, angle and dihedral (“bad”) contributions; these reflect the differences in sugar pucker between A and B helices, and again favor the A form helix in RNA and the B form helix for DNA. (Each of these contributions is discussed in more detail in the original paper.)

The fifth row gives an estimate of the configurational entropy difference between the A and B form helices. As mentioned above, this can be quite a difficult quantity to estimate. The numbers given in Fig. 2 are based on normal mode analyses and suggest that the A and B form helices are about equally stiff. The “total” values in red are the sum of the rows above them, and indicate the the B helix is preferred by DNA, and the A helix by RNA, by roughly 1 to 2 kcal/mol per base-pair. The extra effects arising from added salt (shown in the last two lines of the table) further stabilize the A-form helix relative to B, since the former has more unfavorable bare electrostatic interactions to be screened; this effect quickly saturates, so that 0.1 M added salt is nearly as effective in screening as is 1.0 M added salt. For this sequence, the B-form helix in DNA is favored one at all salt concentrations, but other, more “A-philic” sequences are predicted to undergo a B to A transition at high salt, in rough agreement with experiment.⁶⁴

In cases where specific hydration or ion interaction is important for characterizing molecular interaction, for example with minor groove binders to DNA where hydration in the grooves facilitates the recognition or with G-quadruplex DNA where Na⁺ or K⁺ ions are integral parts of the structure, explicit waters or ions can be included in the MM-PBSA analysis.⁷¹ Critical for proper inclusion of the ions are choosing effective radii that lead to accurate free energies of hydration (Born energies) for the ions. The earlier work cited that began to expose problems with the backbone distributions investigated the minor groove binding drug DAPI bound to different DNA sequences (binding in either the AATT region or shifted by one base pair as seen in experiments). To get reliable estimates of the relative binding affinities (in simulations that were on the ~10 ns timescale (which is before significant disruption in the backbone distributions occurred, which is typically on the 10–50 ns timescale) required inclusion of a small number of bound explicit waters in the MM-PBSA analysis.¹² Related work showed that including ions in G-quadruplexes or potential

intermediates on the folding pathways was absolutely necessary to allow for meaningful energetic comparisons.⁷²

In spite of many promising results, the general performance of the MM-PBSA model in studying ligand binding to proteins and nucleic acids has not been good enough for straightforward use as a part of a ligand design strategy. Beyond the obvious (but persistent) difficulty in obtaining good binding poses in the absence of experiment, a part of this probably arises from the large cancellation noted above between direct electrostatic effects and solvent screening; unless this balance is accurate (for many different conformations) trends in binding energies involving only a few kilocalories per mole may be lost. Continuum solvent models almost certainly fail as well to capture localized hydration effects or trends in small cavities where the size of an individual water molecule is key to understanding what “fits” and what doesn’t. Beyond this, the need to estimate changes in configurational entropy is a weakness, since there is no general procedure for making such calculations.

An early, but still relevant, example considered the binding of caffeine-like molecules to an RNA aptamer, where observed affinities cover a range of 6 kcal/mol.⁷³ Here the ligand molecules are quite rigid (and similar to each other), and the binding pocket is well-defined, characterized by hydrophobic intercalation between bases and water-accessible regions along the edge of the binding site. Relative binding energies were faithfully reproduced by explicit solvent free energy (thermodynamic integration) techniques, but MM-PBSA analyses showed errors in relative binding affinities of up to 3 kcal/mol, and which proved very hard to rationalize, even knowing the correct (TI or experimental) result. Other studies have been more positive,^{65–70} but it seems to be hard to predict in advance what level of accuracy to expect.

5 Where are we going?

Efforts by our groups and others aim to further validate, assess, and improve the nucleic acid force fields in explicit solvent and to also develop better methods for adequately sampling the conformational distributions of nucleic acids. This is being facilitated by performing many MD simulations with different force fields on common nucleic acid motifs and duplex structures. Key applications of the methods center on better understanding RNA-ligand interaction, providing better methods to refine NMR structures, and ultimately being able to model the subtle influence of the environment on nucleic acid structure, for example with riboswitch folding induced by ligand binding. Areas where active and promising research is underway includes the development of constant pH methods in biomolecular simulation and development of enhanced replica-exchange capabilities beyond temperature, such as to exchange Hamiltonians. These alternate Hamiltonian models could be subsets of NMR restraint data, different force fields, different biasing potentials, or effectively many possible changes in the underlying energetic representation. Further force field changes are also under investigation to improve sugar puckering, the conformational distributions of dinucleotides, and also to improve the geometries of the nucleotides. As discussed above, the use of implicit solvent models for nucleic acids has a mixed history, at least in our hands. Duplexes of both DNA and RNA appear to be near the edge of stability, so that minor changes (*e.g.* in the parameterization of GB models) can lead to unstable simulations, with significant conversion to single-stranded configurations. There are generalized Born models that give good results for DNA duplexes,⁵⁵ and these can be scaled up to calculations on hundreds of base pairs (as discussed above), seemingly with good results. RNA helices are less well-described (with wide major grooves), and the configurations of single-stranded regions (which are more prevalent in RNA than DNA) are a problem spot for all types of simulations.³⁵

It is likely that real progress will require fundamentally different types of implicit solvent models, that go beyond the continuum dielectric approximation. One promising direction uses integral equation methods, which averages over a thermal ensemble of solvent configurations, but retains the molecular character of the water and ions surrounding nucleic acids.^{74–76} This method is not nearly as fast as the generalized Born models, but appears to give good results for ion distributions around DNA and RNA, and to fix the problems with the major groove seen in GB or PB calculations (unpublished results).

Because of our background, we have discussed here in detail only nucleic acid simulations using the Amber force fields. This neglect of other work, notably from the CHARMM^{77–79} and GROMOS^{80,81} communities, is a genuine limitation, but it is fair to say that the Amber force fields have been more widely tested, for example by the “ABC” international consortium.^{16–18} Only fairly limited testing on DNA duplex structure has been performed with the GROMOS force fields, which tend to favor A-DNA geometries.⁸⁰ More extensive tests of the CHARMM force fields show a -reasonable representation of DNA duplex structure, albeit with somewhat diminished minor groove narrowing in A-rich regions. Representation of RNA structure has also improved, however RNA structure does not always appear well-maintained in longer simulations with even the newer CHARMM force fields.⁸² The work discussed here has focused on efforts pushing the AMBER force fields to the limit through broad exploration of many nucleic acid models under a variety of conditions and time scales.

In this account, we have also paid considerable attention to problems that have been encountered, especially in simulations on microsecond and longer time scales. This may give an unduly negative impression of the status of MD simulations in this area. Quite to the contrary, with current (fixed-charge) force fields, a great many structural and energetic features of nucleic acid duplexes, including details of sequence dependence, are well-reproduced in modern simulations.^{16–18} Such simulations can fold up and reproduce the fine structure of RNA tetraloops and also refine better NMR structures of nucleic acids.^{83,84} Simulations of single-stranded regions are less predictive, but still capture much of the structural behavior seen in crystal and NMR studies; it can be argued that remaining problems involve small imbalances, for example in the importance of base-stacking interactions versus exposure to solvent. Much practical interest in this field involves complexes of DNA or RNA with proteins; here there are more contacts and structural restraints and one often finds (although we have not reviewed this here) structural results that are in faithful agreement with observation. Challenges for the future include understanding the effects of the ionic atmosphere (particularly for divalent ions like Mg^{2+}), and developing quantitative estimates for the energetics of non-covalent interactions with ligands or proteins. Studies of this sort will continue to build on the foundation of molecular dynamics simulations of the sort described here.

Acknowledgments

This research was supported by NIH grants GM57513, GM103297 and GM81411. The 50th anniversary of Biopolymers coincides with the 25th year that one of us (DAC) has served on its editorial board, and he takes this opportunity to thank other members of the editorial board for many enjoyable interactions. We also thank Don Bashford, Tyler Luchko, InSuk Joung, Niel Henriksen, Rodrigo Galindo-Murillo and Christina Bergonzo for many useful discussions.

References

1. Cheatham TE III, Kollman PA. *Annu Rev Phys Chem.* 2000; 51:435–472. [PubMed: 11031289]
2. Beveridge DL, McConnell KJ. *Curr Opin Struct Biol.* 2000; 10:182–196. [PubMed: 10753816]
3. Cheatham TE III, Young MA. *Biopolymers (Nucl Acid Sci).* 2001; 56:232–256.

4. Giudice E, Lavery R. *Acc Chem Res.* 2002; 35:350–357. [PubMed: 12069619]
5. Orozco M, Perez A, Noy A, Javier Luque F. *Chem Soc Rev.* 2003; 32:350–364. [PubMed: 14671790]
6. Cheatham TE III. *Curr Opin Struct Biol.* 2004; 14:360–367. [PubMed: 15193317]
7. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. *J Chem Phys.* 1995; 103:8577–8593.
8. Sagui C, Darden TA. *Annu Rev Biophys Biomol Struct.* 1999; 28:155–179. [PubMed: 10410799]
9. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman P. *Comp Phys Commun.* 1995; 91:1–41.
10. Cheatham TE III, Cieplak P, Kollman PA. *J Biomol Struct Dyn.* 1999; 16:845–862. [PubMed: 10217454]
11. Wang B, Baker J, Pulay P. *Phys Chem Chem Phys.* 2000; 2:2131–2135.
12. Spackova N, Cheatham TE, Ryjacek F, Lankas F, vanMeervelt L, Hobza P, Sponer J. *J Am Chem Soc.* 2003; 125:1759–1769. [PubMed: 12580601]
13. Varnai P, Djuranovic D, Lavery R, Hartmann B. *Nucl Acids Res.* 2002; 30:5398–5406. [PubMed: 12490708]
14. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. *Biophys J.* 2007; 92:3817–3829. [PubMed: 17351000]
15. Svozil D, Sponer JE, Marchan I, Perez A, Cheatham TE, Forti F, Luque FJ, Orozco M, Sponer J. *J Phys Chem B.* 2008; 112:8188–8197. [PubMed: 18558755]
16. Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE, Dixit SB, Giudice E, Lankas F, Lavery R, Maddocks JH. *Biophys J.* 2004; 87:3799–3813. [PubMed: 15326025]
17. Dixit SB, Beveridge DL, Case DA, Cheatham TE III, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Sklenar H, Thayer KM, Varma P. *Biophys J.* 2005; 89:3721–3740. [PubMed: 16169978]
18. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T III, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Spackova N, Sponer J. 2010; 38:299–313.
19. Beveridge DL, Cheatham TE III, Mezei M. *J Bioscience.* 2012; 37:379–397.
20. Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE, Sponer J, Otyepka M. *J Chem Theor Comput.* 2010; 6:3836–3849.
21. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE, Jurecka P. *J Chem Theory Comput.* 2011; 7:2886–2902. [PubMed: 21921995]
22. Yildirim I, Stern HA, Kennedy SD, Tubbs JD, Turner DH. *J Chem Theory Comput.* 2010; 6:1520–1531. [PubMed: 20463845]
23. Yildirim I, Stern HA, Tubbs JD, Kennedy SD, Turner DH. *J Phys Chem B.* 2011; 115:9261–9270. [PubMed: 21721539]
24. Yildirim I, Kennedy SD, Stern HA, Hart JM, Kierzek R, Turner DH. *J Chem Theory Comput.* 2012; 8:172–181. [PubMed: 22249447]
25. Krepl M, Zgarbova M, Stadlbauer P, Otyepka M, Banas P, Koca J, Cheatham TE III, Sponer J. *J Chem Theory Comp.* 2012 in press.
26. Zgarbova M, Javier Luque F, Sponer J, Cheatham TE III, Otyepka M, Jurecka P. *J Chem Theory Comput.* 2013 in press.
27. Auffinger P, Cheatham TE III, Vaiana AC. *J Chem Theory Comput.* 2007; 3:1851–1859.
28. Joung IS, Cheatham TE III. *J Phys Chem B.* 2008; 112:9020–9041. [PubMed: 18593145]
29. Joung IS, Cheatham TE III. *J Phys Chem B.* 2009; 113:13279–13290. [PubMed: 19757835]
30. Yoo J, Aksimentiev A. *J Phys Chem Lett.* 2012; 3:45–50.
31. Allnér O, Nilsson L, Villa A. *J Chem Theory Comput.* 2012; 8:1493–1502.
32. Besseova I, Banas P, Kuhrova P, Kosinova M, Otyepka M, Sponer J. *J Phys Chem B.* 2012; 116:9899–9916. [PubMed: 22809319]
33. Kuhrova P, Banas P, Best RB, Sponer J, Otyepka M. *J Chem Theory Comput.* 2013; 9:2115–2125.

34. Tubbs JD, Condon DE, Kennedy SD, Hauser M, BEvilacqua PC, Turner DH. *Biochemistry*. 2013;996–1010. [PubMed: 23286901]
35. Henriksen NM, Roe DR, Cheatham TE III. *J Phys Chem B*. 2013; 117:4014–4027. [PubMed: 23477537]
36. McQuarrie, DA. *Statistical Mechanics*. Harper and Row; New York: 1976.
37. Anderson CF, Record MT Jr. *Annu Rev Phys Chem*. 1995; 46:657–700. [PubMed: 7495482]
38. Warwicker J, Watson HC. *J Mol Biol*. 1982; 157:671–679. [PubMed: 6288964]
39. Sharp KA, Honig B. *Annu Rev Biophys Biophys Chem*. 1990; 19:301–332. [PubMed: 2194479]
40. Sharp KA, Honig B. *J Phys Chem*. 1990; 94:7684–7692.
41. Honig B, Nicholls A. *Science*. 1995; 268:1144–1149. [PubMed: 7761829]
42. Baker NA. *Meth Enzymol*. 2004; 383:94–118. [PubMed: 15063648]
43. Baker, NA.; Bashford, D.; Case, DA. *New Algorithms for Macromolecular Simulation*. Leimkuhler, B.; Chipot, C.; Elber, R.; Laaksonen, A.; Mark, A.; Schlick, T.; Schuette, C.; Skeel, R., editors. Springer-Verlag; 2006. p. 263-295.
44. Gilson MK, Davis ME, Luty BA, McCammon JA. *J Phys Chem*. 1993; 97:3591–3600.
45. Jean-Charles A, Nicholls A, Sharp K, Honig B, Tempczyk A, Henderson TF, Still WC. *J Am Chem Soc*. 1991; 113:1454–1455.
46. Bashford D, Case DA. *Annu Rev Phys Chem*. 2000; 51:129–152. [PubMed: 11031278]
47. Brown RA, Case DA. *J Comput Chem*. 2006; 27:1662–1675. [PubMed: 16900491]
48. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL. *J Comput Chem*. 2004; 25:265–284. [PubMed: 14648625]
49. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A. *J Chem Theory Comput*. 2007; 3:156–169. [PubMed: 21072141]
50. Onufriev, A. *Modeling Solvent Environments*. In: Feig, M., editor. *Applications to Simulations of Biomolecules*. Wiley-VCH; Weinheim: 2010. p. 127-165.
51. Srinivasan J, Trevathan MW, Beroza P, Case DA. *Theor Chem Acc*. 1999; 101:426–434.
52. Lange AW, Herbert JM. *J Chem Theory Comput*. 2012; 8:4381–4392.
53. Tsui V, Case DA. *J Am Chem Soc*. 2000; 122:2489–2498.
54. Tsui V, Case DA. *Biopolymers (Nucl Acid Sci)*. 2001; 56:275–291.
55. Gaillard T, Case DA. *J Chem Theory Comput*. 2011; 7:3181–3198. [PubMed: 22043178]
56. Tolbert BS, Miyazaki Y, Barton S, Kinde B, Starck P, Singh R, Bax A, Case DA, Summers MF. *J Biomol NMR*. 2010; 47:205–219. [PubMed: 20549304]
57. Bomble Y, Case DA. *Biopolymers*. 2008; 89:722–731. [PubMed: 18412139]
58. Lin D, Matsumoto A, G N. *J Chem Phys*. 1997; 107:3684–3690.
59. Matsumoto A, G N. *J Chem Phys*. 1999; 110:11070–11075.
60. Matsumoto A, Olson WK. *Biophys J*. 2002; 83:22–41. [PubMed: 12080098]
61. Matsumoto A, Tobias I, Olson WK. *J Chem Theory Comput*. 2005; 1:117–129.
62. Matsumoto A, Tobias I, Olson WK. *J Chem Theory Comput*. 2005; 1:130–142.
63. Matsumoto, A.; Olson, WK. *Normal Mode Analysis: Theory and Applications to Biological and Chemical*. Cui, Q.; Bahar, I., editors. Chapman and Hall/CRC; Boca Raton, FL: 2006. p. 187-211.
64. Srinivasan J, Cheatham TE III, Cieplak P, Kollman P, Case DA. *J Am Chem Soc*. 1998; 120:9401–9409.
65. Kollman PA, 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 DA, Cheatham TE III. *Accts Chem Res*. 2000; 33:889–897.
66. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M. *J Med Chem*. 2005; 58:4040–4048. [PubMed: 15943477]
67. Rastelli G, Del Rio A, Degliesposti G, Sgobba M. *J Comput Chem*. 2010; 31:797–810. [PubMed: 19569205]
68. Brice AR, Dominy BN. *J Comput Chem*. 2011; 32:1431–1440. [PubMed: 21284003]
69. Homeyer N, Gohlke H. *Molec Inform*. 2012; 31:114–122.

70. Hou T, Wang J, Li Y, Wang W. *J Chem Inf Model*. 2012; 24:69–82.
71. Tsui V, Case DA. *J Phys Chem B*. 2001; 105:11314–11325.
72. Stefl R, Cheatham TE III, Sapckova N, Fadrna E, Berger I, Koca J, Sponer J. *Biophys J*. 2003; 85:1787–1804. [PubMed: 12944293]
73. Gouda H, Kuntz ID, Case DA, Kollman PA. *Biopolymers*. 2003; 68:16–34. [PubMed: 12579577]
74. Kovalenko, A. *Molecular Theory of Solvation*. Hirata, F., editor. Kluwer Academic Publishers; Dordrecht: 2003. p. 169-275.
75. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J, Kovalenko A. *J Chem Theory Comput*. 2010; 6:607–624. [PubMed: 20440377]
76. Luchko, T.; Joung, IS.; Case, DA. *Innovations in Biomolecular Modeling and Simulation*. 1. Schlick, T., editor. Royal Society of Chemistry; London: 2012. p. 51-86.
77. MacKerell AD Jr, Banavali N, Foloppe N. *Biopolymers*. 2000; 56:257–265. [PubMed: 11754339]
78. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD Jr. *J Chem Theory Comput*. 2012; 8:348–362. [PubMed: 22368531]
79. Denning EJ, Priyakumar UD, Nilsson L, MacKerell AD Jr. *J Comput Chem*. 2011; 32:1929–1943. [PubMed: 21469161]
80. Soares TA, Hunenberger PH, Kastenholtz MA, Krautler V, Lenz T, Lins RD, Oostenbrink C, Van Gunsteren WF. *J Comput Chem*. 2005; 26:725–737. [PubMed: 15770662]
81. Ricci GG, de Andrade ASC, Mottin M, Netz PA. *J Phys Chem B*. 2010; 114:9882–9893. [PubMed: 20614923]
82. Beššeová I, Banáš P, Kührová P, Košinová P, Otyepka M, Šponer J. *J Phys Chem B*. 2012; 116:9899–9916. [PubMed: 22809319]
83. Kührová P, Banáš P, Best RB, Šponer J, Otyepka M. *J Chem Theory Comput*. 2013; 9:2115–2125.
84. Henriksen NM, Davis DR, Cheatham TE 3rd. *J Biomol NMR*. 2012; 53:321–339. [PubMed: 22714631]

Figure 1.

Low-frequency—in and out of plane bending—modes of circular DNA. The arrows define the main motions associated with the corresponding vibration. Bending motions corresponding to $n=0$ or $n=1$ do not exist in the analytical model, so the first allowed bending comes for $n=2$.

Figure 2.

MM-PBSA analysis of the A to B form transition in d(CCAACGTTGG)². Each column shows values for the B - A energy difference, in kcal/mol: a negative value (as for the total in red of -21.0 for DNA) indicates that the B helix is more stable than the A helix. Values are averages of 100 snapshots from each of four MD simulations (DNA and RNA, each in the A-form or B-form helix.)⁶⁴ Values up to and including the red “total” line refer to zero added salt; values in the last two rows give “total” numbers for 0.1 M or 1.0 M added salt.