Published in final edited form as:

Wiley Interdiscip Rev RNA. 2017 September ; 8(5): . doi:10.1002/wrna.1422.

Physics-Based All-Atom Modeling of RNA Energetics and Structure

Louis G. Smith¹, Jianbo Zhao², David H. Mathews³, and Douglas H. Turner⁴

¹University of Rochester School of Medicine and Dentistry Biochemistry and Biophysics

²University of Rochester Chemistry

³University of Rochester School of Medicine and Dentistry Biochemistry and Biophysics

⁴University of Rochester Chemistry

Abstract

The Potential, U, is Essential: The many conformations from a simulation produce a free energy surface. The surface is realistic, if the force field, FF, is. The database of RNA sequences is exploding, but knowledge of energetics, structures, and dynamics lags behind. All-atom computational methods, such as molecular dynamics, hold promise for closing this gap. New algorithms and faster computers have accelerated progress in improving the reliability and accuracy of predictions. Currently, the methods can facilitate refinement of experimentally determined NMR and x-ray structures, but are less reliable for predictions based only on sequence. Much remains to be discovered, however, about the many molecular interactions driving RNA folding and the best way to approximate them quantitatively. The large number of parameters required means that a wide variety of experimental results will be required to benchmark force fields and different approaches. As computational methods become more reliable and accessible, they will be used by an increasing number of biologists, much as x-ray crystallography has expanded. Thus, many fundamental physical principles underlying the computational methods are described. This review presents a summary of the current state of molecular dynamics as applied to RNA. It is designed to be helpful to students, postdoctoral fellows, and faculty who are considering or starting computational studies of RNA.

Keywords

Structure Prediction; Force Fields; AMBER; Molecular Dynamics; Potentials; Thermodynamics

Introduction

The Encyclopedia of DNA Elements (ENCODE) project, [1] revealed that about 90% of human DNA sequence is transcribed into RNA, but less than 5% of the DNA codes for protein. According to the RNACentral database, [2] over 10 million RNA sequences are

Correspondence to: Jianbo Zhao.

¹These authors contributed equally to this work.

available in 2016. The structures and functions of many of these RNAs are not known. In principle, understanding the physics of RNA interactions would allow translation of sequence into structure and facilitate discovery of functions.

1.1 RNA Structure

RNA structure typically has a hierarchical nature because secondary structure is generally more thermally stable than tertiary structure and often forms on a faster timescale. Therefore, the energetic contributions of secondary and tertiary structures can be treated separately (Figure 1). Secondary structure can be rapidly determined with a combination of sequence comparison, prediction of free energy change for folding, and experimental measurements providing structural restraints [7, 8, 9, 10, 11, 12, 13, 14]

Determination of 3D structures of RNA is difficult and expensive. Thus a large number of important RNA structures remain unsolved. To date, experimental methods such as X-ray crystallography and NMR provide the most reliable determinations of atomic 3D structure. As of 2016, however, only 7% of the PDB's X-ray and NMR structures have RNA.[15] Moreover, crystal and NMR structures report the most populated state or an averaged structure, which are single points on the energy landscape; RNA behavior and function is often dependent on dynamics or the probabilities of alternate conformations in the heterogeneous ensemble on the landscape.[16] For these reasons, computational methods provide an attractive approach to determine or facilitate determination of RNA structures and dynamics.

Physics-based approaches provide particularly general methods for predicting RNA structure and dynamics. Their ground-truth is the physics of molecular interactions, rather than analogies to previously determined structures. Thus, they can be applied to modified RNAs, including those with unnatural modifications developed for synthetic biology.[17, 18, 19, 20, 21, 22, 23] Physics-based methods predict a molecule's thermodynamic ensemble, which is the most important determinant of its function.

This review summarizes basic physical principles for energetic modeling of RNA, reviews applications to predict properties of RNA, and identifies some challenges to providing accurate predictions.

2 Force Fields

In principle, *ab-initio* quantum mechanics (QM) is able to describe all interactions of RNA. Current computers and algorithms, however, do not have the power required to perform simulations using QM-derived energies on even the smallest biologically important RNA molecules.[24] Individual components of RNA, such as base-base interactions, can be studied in vacuum and/or implicit solvent by perturbation theory with large basis sets [25, 26] or highly electron correlated methods such as CCSD(T).[27] It has also been possible to test a broad array of QM methods diverse in computational expense against costly computations with a CCS(T) localized approximation on 46 conformations of the dinucleotide UpU.[28] Comparisons were also made to the AMBER force field. Atomistic

descriptions of complete RNAs in liquid environments is currently beyond the scope of these calculations.

Molecular Mechanics (MM) is a less computationally expensive alternative to QM. In MM, molecular interactions are approximated by force fields.¹ The force field for a molecular system is a potential that mimics the behavior of the molecule when evolved using classical mechanics. If the potential is sampled using discrete dynamics equations then this is called Molecular Dynamics (MD), which is discussed in section 3. This is done by mimicing the average potential energies inherent to the quantum mechanics governing these systems.[29] The realization that bonded atoms tend to have similar characteristics like vibrational frequencies (within several wavenumbers) and bond lengths (within hundredths of an Angstrom) in different molecules led to the development of simple functions with transferability between molecules.[30, 31] Force fields for biopolymers typically express energies as functions of pairwise internal and three-dimensional coordinates. This is done for economy and simplicity of parameterization, which are both important considerations when trying to simulate a system with tens or possibly hundreds of thousands of atoms. It's worth mentioning that some of the many body effects of the liquid environment are baked into certain parameters to be discussed below, so these effects are not entirely absent from simulations done with pairwise additive force fields.[32] A generic potential is:

$$U_{ff} = U_{stretch} + U_{bend} + U_{dihedral} + U_{VDW} + U_{electrostatic} \quad (1)$$

and a particular one is the AMBER force field, which is typical of most pairwise additive force fields used for MD:[33]

$$U_{AMBER} = \sum_{a \in \{bonds\}} \frac{1}{2} k_a (r_a - r_{0,a})^2 + \sum_{b \in \{angles\}} \frac{1}{2} k_b (\theta_b - \theta_{0,b})^2 + \sum_{c \in \{dihedrals\}} \sum_n \frac{V_{n,c}}{2} (1 + \cos(n\phi_c - \gamma_c)) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right) \quad (2)$$

Consideration of many-body effects would complicate these equations by adding cross terms.[30] Definition and development of each term is discussed in the following sections. For more detail the reader is directed to references 30 and 32.

¹**Scales of Economy** Molecular mechanics (MM) is typically based on a pairwise potential, and is often evolved using discrete integrators developed from kinematics. As such, the number of interactions that must be calculated at each step is at worst proportional to N^2 and is linear in memory use where N is the number of bodies in the calculation. In molecular mechanics, part of the approximation is that the bodies are atom centers. By contrast, uncorrelated *ab initio* QM, such as the Hartree-Fock method and methods that are implicitly correlated like Density Functional Theory (DFT), scales at best proportional to N^3 , with memory scaling proportional to N^2 . For QM methods the N is instead the number of electrons in the system. To get a sense of how different this is, consider that for MM two acetylene molecules in vacuum would have eight bodies, not all of which “see each other” with respect to the potential, whereas for QM this system would have twenty eight, all of which interact. Now make the same consideration for a pair of backbone phosphates (PO 4).

2.1 Bond and Angle Potentials

Early work that resulted in a predictive force field for bond energies came from Lifson and Warshel, and was based on the “conjecture that a limited set of elementary energy functions is associated with these bonds, such that the energy of any alkane molecule is obtained as a sum of such elementary functions.”[31] The idea is straightforward; there is empirical behavior that internal degrees of freedom mostly exhibit on average. This behavior can be specified as a Taylor expansion of the potential energy surface in question.

$$U(r) = U(r_0) + \sum_a \left. \frac{\partial U}{\partial r_a} \right|_{r_0} (r_a - r_{0,a}) + \frac{1}{2} \sum_a \left. \frac{\partial^2 U}{\partial r_a^2} \right|_{r_0} (r_a - r_{0,a})^2 + \dots \quad (3)$$

Here the function, $U(r)$, is the energy resultant from the state of the bond, r_a is the length of bond a , with $r_{0,a}$ the equilibrium length. Note that to include cross-terms in this series, one would consider a multidimensional Taylor expansion instead, where the mixed partials would be with respect to separate atoms.

For most organic molecules near room temperature, these expansions can be sensibly truncated at second order. These second order expansions are often accurate up to energies around 9.5 kcal/mol above the equilibrium conformation, which conservatively provides geometric accuracy in a room temperature simulation of a biomolecule.[30] In eq. 3, the term containing the first derivative is by definition zero if r_0 is an equilibrium value. Since the zero of potential energy, U_{r_0} , is arbitrary it is omitted, yielding a simple harmonic function for the internal coordinate.

$$U_{stretch,a} = \frac{1}{2} \left. \frac{\partial^2 U}{\partial r_a^2} \right|_{r_{0,a}} (r_a - r_{0,a})^2 = \frac{1}{2} k_a (r_a - r_{0,a})^2 \quad (4)$$

Here, the partial derivative of the potential with respect to r_a —which, when evaluated at $r_{0,a}$ is a constant—is renamed as k_a to show its relationship to the harmonic approximation. The k_a can be thought of as the force constant for the spring. Summing over all bonds gives the total bond stretch energy for a given conformation (compare eq. 5 to the first term in eq. 2).

$$U_{stretch} = \sum_{a \in \{bonds\}} U_{stretch,a} = \sum_{a \in \{bonds\}} \frac{1}{2} k_a (r_a - r_{0,a})^2 \quad (5)$$

Because k_a corresponds to a force constant in experimental measurements of vibratory modes, and because these degrees of freedom are stiff relative to many involved in macromolecular conformational dynamics, they can be transferred from experimental measurements on small molecules. All that is needed is an electronically similar functional moiety in the small molecule, an approximation inherent in the notion of functional groups.

Of course, mistaking the functional group in the parameterized molecule invalidates this transference, so these parameters should not rise above scrutiny.[34, 35, 36] Note, though, that while transferability is a major benefit of the physical approach, fitting a bond directly from the vibrational spectrum of anything other than a simple molecule is impossible; the spectra of large molecules are too complex to accurately fit to specific bond distances and stiffnesses. Hence for RNA (and other large molecules) these parameters are taken from those fit to the spectra of analogous moieties on small molecules. Distances could be taken from more complete assemblies of these moieties in crystallographic databases, especially those for small and medium-sized molecules, but this has not been done for RNA yet.

Angle distortion is described in the same way as bond distortion (see second term in eq. 1 and second series in eq. 2); if the angle, θ_b , between two bonds on the same atom center in a molecule rests at a natural or equilibrium angle, $\theta_{0,b}$, then a truncated Taylor expansion of the potential energy surface gives a harmonic function of angle that is accurate for small deviations from the rest angle. This works because angle distortion is also stiff when compared to the average energy a degree of freedom might have at room temperature. Equilibrium angles are found—often by looking at sub-Angstrom resolution small molecule crystal structures—then angle distortion force constants are obtained from spectroscopic measurements, where the ‘bend’ modes can be readily distinguished from the ‘stretch’ modes in analogous small molecules.[37, 38]

2.2 Dihedral Potentials

Unlike the previous terms, dihedral potentials have no obvious direct physical transference from experimental data. Most molecules have energetic preferences for the relative orientation of any two bonds connected by a third. The large number of rotatable bonds per residue in the RNA backbone lead to a large number of such preferences—combinations of dihedral basins sometimes called suites.[39] This dependence is expressed as a truncated Fourier series expansion—ideal for periodic functions—of the projective angle between the two bonds:

$$U_{dihedral,c} = \sum_n \frac{V_{n,c}}{2} (1 + \cos(n\phi_c - \gamma_c)) \quad (6)$$

For each dihedral, there is a period, n , a phase shift, γ , and a coefficient to bring the barrier heights to the correct scale, $V_{n,c}$. The period is sometimes called the order of the dihedral, and is usually a free parameter in the fitting process.

$$U_{dihedrals} = \sum_{c \in \{dihedrals\}} U_{dihedral,c} = \sum_{c \in \{dihedrals\}} \sum_n \frac{V_{n,c}}{2} (1 + \cos(n\phi_c - \gamma_c)) \quad (7)$$

Summing over all dihedrals as in eq. 7 gives the dihedral energy for a conformation (the third term in eq. 2). These parameters are assigned to match barriers calculated using QM or, when possible, to match those inferred from experiment. Dihedrals can be seen as error

correction terms and are usually fit last in force field development. Force field tuning is most commonly accomplished through dihedral adjustment or refitting.[40, 32, 41, 42, 43]

Most force fields also define dihedrals between atoms not connected in series. Sometimes, these “improper” dihedrals are used to ensure planarity in SP² hybridized systems like RNA bases. Correct inclusion of cross-terms in the angular energy can also result in the correct planarity of these systems.[30] To our knowledge, however, in RNAs and other biopolymers this behavior is always modeled with improper dihedrals.

2.3 van der Waals Potentials

Structured RNA is extensively stacked. In water, nucleobases will stack with one another like coins in a roll instead of interacting edge-on-edge—with organic aprotic solvents reversing this trend.[44, 45, 46] Stacking may result from several physical phenomena: The dispersion effect gives a favorable interaction between electrons in the aromatic systems.[47, 48] Bases have functional groups with partial charges, and overlap of opposite charges, e.g. an amino group over a carbonyl also favors stacking[49] Actual overlap between orbital charge distributions may occur, causing a charge penetration effect[29] that has been shown to be favorable.[50, 51] Stacking avoids the loss of a water-water hydrogen bond that occurs adjacent to the flat surface of the base.[52, 53, 54]

The latter effect differs from the entropy-driven classical hydrophobic effect described for more spherical solutes.[55] With nucleobases, water is oriented near solute as in the hydrophobic effect, but because of the planarity of the base a water directly adjacent to the aromatic ring cannot orient to provide a good water-water hydrogen bond. Instead, one water over this surface will reorient to form a weaker hydrogen-pi interaction with the aromatic system, optimally aligning the remaining water-water hydrogen bonds present. This rationalizes the observation that stacking is enthalpically driven, since the loss of the water-water hydrogen bond would be enthalpic.[56, 54]

In contrast to bond distances and angles, the energetics of nucleobase stacking depend on non-bonded interactions. Such interactions are present between all atoms in a molecule, but are weaker than bonded interactions. These interactions can be broken up into long-range electrostatic effects and close-range van der Waals (VDW) effects. This part of the force field is much more challenging to fit because the effects are more subtle.

The VDW interaction is driven by the combination of dispersion correlation and exchange repulsion that causes molecules to adhere to and bounce off one another, respectively. The dispersion-induced interaction can be seen as a multipole expansion of the induced electronic structure from two proximal molecules, which can be sensibly truncated at dipole order. Thus the approximate VDW energy is a function of the distance between atoms *i* and *j*, R_{ij} , and the attractive term should be proportional to R_{ij}^{-6} , a common distance dependence for dipole dipole effects,[30, 29] leading to an expression like the following:

$$U_{VDW}(R_{ij}) = U_{repulsion}(R_{ij}) - \frac{B_{ij}}{R_{ij}^6} \quad (8)$$

While the attractive term has some physical basis, the exchange repulsion term is murkier. It should be steep, and it would be best if it were cheap to calculate. Thus, squaring the value of R_{ij}^{-6} is a suitable approximation as used in eq 2.

$$U_{VDW} = \sum_i \sum_{j>i} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \quad (9)$$

This expression can be refactored to give physical significance to the coefficients:

$$U_{VDW} = \sum_i \sum_{j>i} \varepsilon_{ij} \left(\left(\frac{R_{0,ij}}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{0,ij}}{R_{ij}} \right)^6 \right) \quad (10)$$

Here ε_{ij} is the energy minimum, and carries units of energy. $R_{0,ij}$ is the position of the minimum, and carries distance units. Equations 9 and 10 are alternative forms of the Lennard-Jones potential, originally developed by John Lennard-Jones to empirically describe the behavior of gases.[57] Eq. 9 is preferred by force field developers because it permits a linear fit of the coefficients, A_{ij} and B_{ij} . Using a function like a Morse potential would be more accurate, especially for modeling the repulsive behavior, but requires the evaluation of one exponential per interaction (and the fitting of three parameters per interaction, rather than only A_{ij} and B_{ij}).[30, 58]

The number of combinations of interactions between functional groups in a biopolymer is enormous. Therefore a formalism is required for generalizing interactions between arbitrary pairs of atoms in the molecule. These schemes are called combination rules. For AMBER, each atom has assigned values for its VDW radius, $R_{0,i}$ and well depth, ε_i .

$$R_{0,ij} = R_{0,i} + R_{0,j} \quad (11)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \quad (12)$$

The minimum for the VDW energy between two atoms is then inferred using eq. 12 [59] and the position as eq. 11 [60]. Though most force fields use these Lorentz-Berthelot combination rules, there are a plethora of others, many emerging from a consensus that equations 11 and 12 are insufficient for liquid simulations.[61, 62] In any case, the parameters to be fit are the atom-associated VDW radii and well-depth.

Various approaches are taken to fit radius and well-depth. A common component of these approaches is adjusting well-depth and radii until macroscopic properties of pure liquids, like the density and solvation entropy, are predicted well. This approach was originally taken

by OPLS[63, 64] and adopted by AMBER because of its successes.[37] CHARMM's approach is to adjust VDW parameters until they correctly reproduce packing geometries and energies of crystal structures as well as ab initio energies of interaction for model compounds.[38, 65] It is important to have parameters developed in a consistent way so that "mixing and matching" should be avoided.

All force fields modify at least some of the VDW parameters associated with polar hydrogens as a substitute for, or sometimes in addition to, an explicit hydrogen bonding potential. AMBER scales its hydrogen VDW radius for polar hydrogens, then lets the point charges (discussed in the section on electrostatics) reflect the variable strengths of the hydrogen bonding interaction.[37] CHARMM tunes the parameters for polar hydrogens by scanning the position of a water molecule, then adjusting VDW parameters until the interaction looks like a water hydrogen bond.[65] As far as we are aware, no currently popular force fields tuned for biopolymers explicitly represent the hydrogen bond.

2.4 Electrostatic Potentials

For polar systems, the calculations done to approximate the electrical field are often the largest part of the calculation, in terms of both expense and energetics; RNA is a polar and charged molecule solvated in a polar, protic solvent (water) and bathed in counterions. Electrostatic terms are therefore quite significant for the accuracy of MM representations of RNA. Moreover, electrostatic effects are long-ranged enough that every atom responds to every other atom's electric field.² Thus accurate mimicry of electrostatics is essential, and computational cost is sensitive to the nature of the approximation.

The approach of biomolecular force fields has almost uniformly been to approximate the electrostatic potential (ESP) of the molecule as a series of fixed partial point charges centered on the atoms of the molecule[66, 67, 68, 63] (but see the concluding remarks). This approximation has limitations; it only represents the average polarization of a molecule, a conformation-dependent property.[69, 70, 71, 72] It also becomes a less accurate approximation for the molecule's effective field as the point evaluated gets closer to the molecule. Like a pointillist painting, what blends and becomes indistinguishable from a uniform charge distribution at a distance separates into individual points of charge up close. This problem is partially mitigated by also having VDW potentials implicitly representing short-range interactions between orbitals. Even outside of this radius, for fitting methods trying to use a QM ESP as an optimization target, the errors tend to be the largest for regions closest to the molecule (see for example reference 73 figures 4–6, and compare gridpoints in the innermost shell with more distant ones).

²**What is Long-Ranged?** Interactions between monopoles (charges) fall off as r^{-1} but a sense of scale is needed. Consider the interaction between two phosphates (negative charges): then $U = Cq_1q_2/er$. Here $q_1 = q_2 = -1$. In the preferred thermochemical units,

$$332 \cdot \frac{kcal}{mol}$$

Coulomb's constant becomes $\frac{332}{mol}$ if distance is provided in number of Å. Assuming unit charges in a vacuum and applying this formula suggests that for the interaction between those charges to drop below 0.1 kcal/mol, they would need to be 3,320 Å apart. In a liquid charge does not penetrate that far, but in explicit solvent simulations this is represented by having all of the charges present and empirically determining the screening through the simulation.

How these sensitive terms are fit is essential to their performance. One approach is to refine the charges in the same way as VDW interactions, by trying to recapitulate thermodynamic properties of liquids.[63] However, if the VDW interactions adequately handle short-ranged artifacts, and if the potential is to be used in an all-atom simulation (where no continuum approximations are used at greater distances) then fitting the charges to a QM-derived ESP in regions of space outside the VDW radius of the molecule also seems appealing.[68] A refined version of this idea is discussed in box 2.4.³

³**To Charge a Model** The full expression for the electrostatic potential of a molecular system, with the i th nucleus at position $\mathbf{R}_i$ and having charge Z_i is:[30, 74]

$$\phi(\mathbf{r}) = \sum_{i \in \{\text{nuclei}\}} \frac{Z_i}{\|\mathbf{r} - \mathbf{R}_i\|} - \int \frac{\Psi^2 \mathbf{r}'}{\|\mathbf{r}' - \mathbf{r}\|} d\mathbf{r}' \quad (13)$$

The charge density, Ψ^2 , can be calculated with QM theory. By applying Coulomb's law to point charges on nuclei, the electrostatic potential can be approximated as:

$$\phi_{\text{points}}(\mathbf{r}) = \sum_{i \in \{\text{atoms}\}} \frac{q_i}{\|\mathbf{r} - \mathbf{R}_i\|} \quad (14)$$

The goal is to minimize the difference between ϕ and ϕ_{points} on a discrete grid around the molecules:

$$\chi_{\text{ESP}}^2(\mathbf{q}) = \sum_{\mathbf{r} \in \{\text{grid points}\}} \left(\phi(\mathbf{r}) - \sum_{i \in \{\text{atoms}\}} \frac{q_i}{\|\mathbf{r} - \mathbf{R}_i\|} \right)^2 \quad (15)$$

Upon differentiation with respect to charge this objective function becomes a system of linear equations that are then optimized. It can be ill conditioned for partial charges that are poorly statistically determined by the grid, for example nonpolar moieties on polar molecules. The remedy for this issue is to hyperbolically restrain the charges to zero:[73]

$$\chi_{\text{Restraint}}^2 = a \sum_{i \in \{\text{atoms}\}} \sqrt{q_i^2 + b^2} - b \quad (16)$$

Here a is a scale factor which defines the asymptotic limits of the restraint strength, and b determines the tightness of the hyperbola around its minimum.

$$\chi^2 = \chi_{\text{ESP}}^2 + \chi_{\text{Restraint}}^2 \quad (17)$$

This approach is further refined if applied across different conformations of a molecule, and by constraining chemically identical charges to be equal using Lagrange multipliers rather than post hoc averaging.[71, 72, 73, 75] These improvements were all used to develop the partial charges for the AMBER model for RNA.[37] However, conformational dependence of charges is still averaged, thus encouraging the tuning of dihedral parameters to resolve conformational preference.[40] For a more detailed derivation of the fitting equations look to reference Bayly 1993 and the appendix in reference Besler 1990.

2.5 Hydrogen Bonding

Hydrogen bonding is a complex range of phenomena resulting from the hydrogen proton being exposed in the context of polar hydrogens.[77] Not all hydrogen bonds are equal. Weak hydrogen bonds are sensibly approximated with electrostatics and van der Waals interactions,[37] but modeling the electron sharing involved in stronger hydrogen bonds is more challenging.[78] These categories are not hard and fast, though approaches to classify them that translate to energetic approximations have been developed.[79] It has been proposed that certain orientations of nucleobases produce resonance-assisted hydrogen bonds, which would lend particular stability to those conformations.[80] Hydrogen bonds in folded RNA can be stronger than those formed between unfolded RNA and water.[81, 82, 83]

2.6 Solvent

2.6.1 Water—Because of the polyanionic nature of RNA molecules, the solvent and counterions interact strongly with the RNA surface; thus they are an integral part of RNA structure.[84] In the crystal structure of an RNA duplex the 2'-OH group increases hydration of the sugar-phosphate backbone relative to DNA, and the bases in the minor groove also have more contacts to water.[85] Some water molecules are predicted to occupy spatially ordered hydration sites near RNA, and they may help stabilize RNA structure by forming hydrogen bond networks in the shallow groove.[86] A simulation study [87] generated an average of 7 water-RNA hydrogen bonds per nucleotide. Specifically, the non-bridging oxygens of the phosphate group, which are the most exposed to solvent, form 2–2.5 hydrogen bonds; O2', O3', O4' and O5' in the ribose group form 2.5 hydrogen bonds in total; N and O moieties in the base also form hydrogen bonds with water. Moreover, the average lifetime of a water-RNA hydrogen bond was predicted to be 1–2 orders of magnitude longer than the lifetime of hydrogen bonds in bulk water. Water molecules located in the hydration layer suffer from boundary restrictions on the hydrogen bond network of water molecules. As a result, the structure and dynamics of these waters behave differently from that of bulk water. Thus, water interactions with RNA may be stronger than interactions with bulk water.

Considering the large number of water molecules in the hydration shell and the long residence time, selection of an explicit water model is crucial in MD simulations. There are many computational models available at different levels of approximation. The most commonly used water models in biomolecular studies are nonpolarizable rigid models with atom-centered point charges, for example TIP3P[88] (Figure 2). Each water model is developed to fit certain experimental parameters, such as the radial distribution function [89] and density. [88] The quality of the model is determined by the ability to reproduce as much as possible various experimental bulk water properties, for instance, self-diffusion coefficient, heat of vaporization, and isobaric heat capacity. Currently, TIP3P [88] and SPC/E models [90] are the most popular due to their low computational cost and relatively high accuracy. The OPC model [91] is able to reproduce several experimental bulk properties more accurately, and is starting to be tested in simulations with RNA.[92] Certain force fields, such as CHARMM's, are tuned to perform with a certain water model, meaning

that even if a model mimics bulk water better, its interactions with RNA parameters tuned to a different model may be less physical.[38, 65]

The choice of explicit water model significantly affects the structure and dynamics of RNA. Simulations carried out with TIP3P and SPC/E models give unfolded and kinked RNA conformations, respectively, of reverse kink-turns.[93] The different behaviors are most likely due to a difference in self-diffusion constant leading to faster unfolding in TIP3P. The water model can also affect conformational ensembles of single stranded RNA tetramers. The OPC model helps penalize overly hydrogen bonded structures, and improves agreement with NMR spectra; but this effect is sequence dependent.[92]

2.6.2 Salt—Ions are required in simulations to neutralize the charge in the box of water containing the RNA. RNA interacts with ions both by attracting clouds of positive ions nonspecifically—these ions remain hydrated—to neutralize charge, and by chelating metal cations in specific binding sites. Some of these binding sites are high affinity and the complexes can have remarkably long half lives.[94] Thus, changes in ionic strength can have marked effects on RNA behavior. Therefore, correct ion-solvent behavior and ion-nucleic acid interactions are essential for predicting realistic ensembles, especially those containing specifically bound positive ions.[95, 96]

Divalent cations clearly affect the ensembles of structured RNAs.[97, 98] Parameterizations of these ions are fraught even in the context of polarizable force fields.[99, 100] For example, the lifetime of RNA-divalent cation interactions can be on the order of milliseconds, possibly longer.[101, 102] This is longer than most MD simulations. While *in-vivo* RNAs nearly always have access to Mg 2+, many useful simulations have not included Mg 2+. A benchmark comparing several models of Mg 2+ bound to two structured RNAs at numerous Mg 2+ sites in multiple 100–200 nanosecond MD simulations to Density Functional Theory QM (DFT) and PDB statistics has been reported.[103] The report suggests that simulations with current Mg 2+ models show context-dependent performance and provides recommendations about which Mg 2+ parameters to use in a given context. DFT does not account for electron correlation explicitly, a concern since many of the binding sites benchmarked involve polarizable groups chelating the Mg 2+.

2.7 Summary of Current Force Fields

The sections above outline a framework for parameterizing and tuning a physics-based model for predicting RNA properties. Because parameterizing a force field depends largely on transferable properties measured in small molecules and on systematized QM calculations, parameters can be developed for essentially any nucleic acid. For example, AMBER and CHARMM have used QM to add parameters for naturally modified nucleic acids.[19, 20] Increases in computing power and sophistication of QM calculations have allowed more accurate modeling of the physics (Figure 3).

Force fields for RNA are less developed than for proteins. As with proteins, bond distances, angles and their energetics are approximated from crystal structures and vibrational spectroscopy. Modeling other components of potential energies of RNA 3D structure, however, remains challenging for several reasons. First, each nucleotide has seven rotatable

torsion angles, including around the glycosidic bond, leading to many degrees of freedom for the RNA backbone.[39] Second, RNAs are largely stabilized by base stacking interactions, which are hard to treat theoretically. Reliable results appear to require the use of highly correlated methods.[28] Third, hydrogen bonds and hydrogen bond networks are essential for the stability of RNA in aqueous solution, but modeling interactions between polar atoms is also challenging. Last, the negatively charged phosphate groups require accurate description of long-range electrostatic interactions and counterions.

2.8 New Fitting Approaches

Recently new approaches have emerged that vary significantly from the aforementioned methods. Three have been applied to fixed point charge force fields for RNAs: Optimal Point Charge (OPC) placement,[104, 91] Targeted Metadynamics (TMD) projection,[105, 106, 107] and multiple linear regression.[108]

2.8.1 Optimal Point Charge Placement—As discussed above, polar compounds are usually modeled as collections of internal degrees of freedom with both VDW and electrostatics terms. The inclusion of electrostatics is expensive, and is therefore often approximated as a collection of fixed point charges centered on the nuclei of the atoms in the molecules. There are two approximations here: the charges do not fluctuate and the charges are directly on the atom centers. Relaxing the second requirement allows for a more exact fitting of the QM Electrostatic Potential (ESP). Performance in water simulations shows promise for applications to other systems, both because the OPC model performs well in AMBER simulations of RNA and because the charge fitting scheme could be applied to RNA moieties that have proven troublesome for current non-bonded parameters.[91, 92] An algorithm has been developed for generating optimal point charges for a general ESP or from a set of point charges.[104]

2.8.2 Targeted Metadynamics Projection—Metadynamics is an enhanced sampling method based on reaction coordinate selection that forces a system to cross barriers by stepwise addition of biasing potentials along the reaction coordinate until the statistical distribution across the coordinate is uniform (e.g. see Figure 4, panels A and B). Unbiasing then gives the free energy curve along the reaction coordinate.[109] The target distribution need not be uniform, however, which opens the possibility of enforcing a target distribution known from data. This Targeted Metadynamics (TMD) allows other aspects of the system to be forced to become consistent with that data.[105] If the reaction coordinate is a degree of freedom parameterized directly by the force field, such as a dihedral, then the weighting necessary to produce that distribution given the rest of the force field can be found using TMD, yielding an empirical potential that models the data characterizing it. This approach has been applied to RNA backbone dihedrals.[107]

2.8.3 Multiple Linear Regression—One of the central challenges of fitting a model with a large number of parameters is that error is rarely sequestered in particular parameters and individual parameters are not independent of each other. This is especially true for parameters that include correction factors and are tuned relative to one another initially. Because dihedrals are often used to compensate for approximations in other parts of the

force field, they experience the most tuning; If one is found to be problematic, e.g. the dihedral,[110, 43] then tuning it can fix some problems but introduce others because the other dihedral parameters were tuned and checked with reference to the old behavior of the adjusted potential.

One way to minimize the problems with stepwise parameterization is to refit such parameters simultaneously to reference data that should inform on all of them. This strategy has been applied by Aytenfisu and coworkers,[108] who simultaneously refit RNA dihedrals to match QM energies. Fitting a large number of parameters requires a large reference database. Thus, a large fitting ensemble of QM energies was generated from dinucleotide conformations in the PDB, augmented with fixed-point QM scans across barriers. Multivariate linear regression was then used to deterministically fit all of the coefficients in the series describing α , β , γ , ϵ , ζ , and, for each base, χ , so that they would not only match their particular profile better, but also be in accordance with each other. While problems remain after applying these parameters, they improve performance for backbone intensive benchmark RNAs like tetramers, notably reducing the prevalence of unrealistic intercalation (see Benchmark section) in these small RNAs. On small base-paired RNAs their performance is similar to the standard force field.

3 Molecular Dynamics

A force field and the laws of physics allow molecular dynamics simulations to predict structures and dynamics of biomolecules. Molecules are given kinetic energy, regulated to a specific temperature, and allowed to evolve in the energy landscape defined by the force field. Simulations must be extended sufficiently to allow adequate sampling of the landscape. This section discusses how the simulations are performed, and how they are interpreted to make connections to macroscopic quantities.

3.1 Integrating the Equations of Motion

In molecular dynamics (MD), the RNA molecule samples conformations along a trajectory, i.e. a path, that is calculated for the atoms as a function of time. To do this, the force field is evaluated at the given conformation, the atoms are moved for a short time step assuming that the potential is constant across that time step, and the process is repeated. This is called integrating the equations of motion.⁴ The time step is governed by the fastest motions in the system. For RNA, they are the vibrations of bonds between hydrogens and heavy atoms, and the longest time step is 1 fs at 300 K. Usually, simulations are constrained along the distance

⁴To model the motion of the atoms, the integrator needs to conserve energy and be reversible with time. To appreciate the subtlety of an integrator, start with a Taylor expansion of position as a function of time:

$$\mathbf{R}(t+\Delta t)=\mathbf{R}(t)+\frac{d\mathbf{R}(t)}{dt} \cdot \Delta t+\frac{1}{2}\frac{d^2\mathbf{R}(t)}{dt^2} \cdot \Delta t^2+O(\Delta t^3) \quad (18)$$

Where $\mathbf{R}(t)$ is the (vector) position of the atoms as a function of time, t is the timestep. The last term is the error term from truncating the expansion, and it is of order Δt^3 . This is the same as writing the Euler equation that is commonly seen in a first physics class:

of these bonds using SHAKE,[122] or related methods.[123] This constraint then allows a timestep of 2 fs at 300 K.

$$\mathbf{R}(t+\Delta t)=\mathbf{R}(t)+\mathbf{v}(t) \cdot \Delta t+\frac{1}{2}\mathbf{a}(t) \cdot \Delta t^2-O(\Delta t^3) \quad (19)$$

where $\mathbf{v}(t)$ is the (vector) velocity and $\mathbf{a}(t)$ is the (vector) acceleration of the atoms. The acceleration is the force divided by the atomic mass, where force is the negative gradient of the potential energy, as determined by the force field:

$$\frac{d^2\mathbf{R}(t)}{dt^2}=\mathbf{a}(t)=\frac{-\nabla U(\mathbf{R})}{m} \quad (20)$$

The velocity estimate in the Euler equation, however, is comparatively crude:

$$\frac{d\mathbf{R}}{dt}=\mathbf{v}(t)=\frac{\mathbf{R}(t)-\mathbf{R}(t-\Delta t)}{\Delta t} \quad (21)$$

i.e. the velocity is estimated from the change in positions from the previous step. In practice, this crude estimate leads to errors that would propagate from step to step, and thus the Euler method cannot be used.[119] To improve the integrator, the next step is to expand the position backwards in time:

$$\mathbf{R}(t-\Delta t)=\mathbf{R}(t)-\frac{d\mathbf{R}(t)}{dt} \cdot \Delta t+\frac{1}{2}\frac{d^2\mathbf{R}(t)}{dt^2} \cdot \Delta t^2-O(\Delta t^3) \quad (22)$$

Adding the forward and backwards expansions (eqs 18 and 22) gives:

$$\mathbf{R}(t+\Delta t)+\mathbf{R}(t-\Delta t)=2\mathbf{R}(t)+\frac{d^2\mathbf{R}(t)}{dt^2}\Delta t^2+O(\Delta t^4) \quad (23)$$

This can be arranged to show:

$$\mathbf{R}(t+\Delta t)=2\mathbf{R}(t)+\mathbf{R}(t-\Delta t)+\frac{d^2\mathbf{R}(t)}{dt^2}\Delta t^2+O(\Delta t^4) \quad (24)$$

Now the positions are propagated without an evaluation of velocity, and the error terms are of the order of Δt^4 . This is called the Verlet algorithm.[120] Velocities can be estimated using:

$$\frac{d\mathbf{R}(t)}{dt}=\mathbf{v}(t)=\frac{\mathbf{R}(t)-\mathbf{R}(t-\Delta t)}{\Delta t} \quad (25)$$

In practice, other integrators, leapfrog and velocity Verlet, which both derive from Verlet, are used because they not only provide good position estimates, but also better estimates of velocities.[121]

3.1.1 Connecting MD to thermodynamics—Much of the power of molecular dynamics derives from its ability to model statistical mechanics from the estimated time evolution of a single molecule. This connection is made using the ergodic hypothesis:

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(t) dt = \frac{1}{N} \sum_{i=1}^N A_i = \langle A \rangle \quad (26)$$

This asserts that with enough time sampling, the time average, $\bar{A}$, of any quantity, A , that can be measured from the simulation, is equal to the ensemble average, $\langle A \rangle$, for some large number, N , of molecules in the ensemble. In other words, the thermodynamic ensemble can be estimated by taking snapshots at regular intervals from the time trajectory. Conditions for the ergodic hypothesis to be true are not generally met because current computer power typically limits molecular dynamics simulations to the microsecond time scale, which is usually not sufficient to find the limit at infinite time. Nevertheless, MD can provide useful approximate answers to carefully constructed questions for RNA systems.

3.2 Methods

3.2.1 Brute force free energies—For an equilibrium between conformations C1 and C2:

The equilibrium constant, K , is:

$$K = \frac{[C_2]}{[C_1]} \quad (28)$$

where the brackets indicate “concentration of.” Assuming the molecules are in a dilute solution, so that the overall volume is constant, then the above equation is equivalent to:

$$K = \frac{N_2}{N_1} \quad (29)$$

where N_1 is the number of molecules in conformation 1 and N_2 is the number of molecules in conformation 2 in the ensemble. If the ergodic hypothesis is true for sampling these two conformations, however, then N_1 is also the number of snapshots in the trajectory that are in conformation 1 and N_2 is the number of snapshots in conformation 2. The equilibrium constant provides the free energy change:

$$\Delta G^\circ = -RT \ln K \quad (30)$$

where R is the gas constant, T is the temperature in kelvins, and ΔG° is the standard Gibbs free energy change, which is applicable for simulations run in constant pressure and temperature.

Determining conformational free energy changes directly from a simulation in this way works as long as there are a number of transitions back and forth between the two conformations. This means the barrier in the free energy landscape cannot be much larger than thermal energy, kT , the product of the Boltzmann constant and the absolute temperature.

3.2.2 Umbrella Sampling—Often one is interested in knowing the free energy change between conformations separated by a barrier larger than kT . Consider the energetic landscape shown in Figure 4A. Because the landscape is being sampled *in-silico*, it could be altered, as in TMD, by adding an energy term that lowers the barriers (Fig. 4B). Then, the simulation would sample between the two conformations, and the effect of the additional energy term could be accounted for in determining the free energy changes. Such amendments to the potential are called biases because they enforce selective sampling of the free energy landscape.

In practice, however, determining the barrier usually requires sampling across the barrier. Because the shape and height of the barrier are unknown, the additional potential to reduce the barrier is also unknown. Although this approach is impractical because of this difficulty, it motivates an approach that is practical. The alternative is to perform a series of independent simulations along the direction of the reaction coordinate, where each simulation is restrained to a position along the reaction coordinate with a harmonic potential (illustrated in Fig. 4C). This approach is called umbrella sampling because each harmonic potential is shaped like an upside down umbrella. For each simulation, it can be shown that the effect of the restraining potential can be removed to find an unbiased free energy difference for that one simulation (Fig. 4D).

To get the overall free energy, the offset between the free energies of adjacent simulations needs to be determined. This is usually accomplished with the weighted histogram analysis method (WHAM), which minimizes the estimated errors in determining the offsets.[124] This can be thought of as “gluing” the free energy curves from adjacent windows together. It is for this reason that good overlap between adjacent windows along the reaction coordinate is essential. A popular program implementing this algorithm is available from Alan Grossfield.[125]

3.2.3 Alchemical Free Energy Change—A second type of free energy calculation predicts which of two structures is more stable in a given context. For example, is it more favorable to be protonated or deprotonated,[126] or is an isoG-isoC pair more stable than a G-C pair (Figure 5) in a given duplex.[112] For these calculations, the change involves altering a chemical structure in a path that is not physically feasible, and therefore it is referred to as alchemy.

One common approach to estimating alchemical free energy changes is to use free energy integration. A parameter, λ , describes the progress along the reaction, where zero is used to indicate the reactant molecule, and one is used to indicate the product molecule. The molecule is then transformed from reactant to product. It is often convenient to use a linear progress formulation, where the potential energy of the system is described as:

$$V(\lambda) = (1-\lambda)V_1 + \lambda V_2 \quad (31)$$

For values of λ greater than 0 and less than 1, the potential energy of the system is a hybrid between the two molecules. One formulation for estimating the free energy change is called thermodynamic integration, where the free energy change is estimated by integrating the partial derivative of the potential with respect to λ , along λ :

$$\Delta G^\circ = \int_{\lambda=0}^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (32)$$

where the angle brackets indicate an ensemble average at a specific λ . A typical way to solve this is to evaluate the derivative with equilibrium simulations at multiple λ , the integral is solved numerically, for example using a Gaussian quadrature.[127, 33]

3.2.4 Convergence—When analyzing MD simulations, an important consideration is whether there is enough sampling to address a quantity of interest. As mentioned above, all-atom MD simulations are currently limited to microseconds of sampling, and therefore the amount of sampling is limited. Acquiring enough sampling is called convergence, i.e. a simulation that has adequate sampling to be ergodic is said to be converged.

Because MD is a conformational sampling method, it is subject to statistical error because the sample size is limited. Larger samples reduce the statistical error, and convergence is a matter of extent. One way to address convergence is to treat it as a statistical problem by quantifying the statistical uncertainty.[128] A common approach to assess convergence is to ascertain the statistical errors based on the set of conformations sampled from a single trajectory. In this approach, it is important to realize that conformations from snapshot to snapshot along the trajectory are correlated and therefore the sample size is generally smaller than the number of trajectory snapshots. Block averaging is one procedure to assess the sample size.[128, 129] Another approach is to assess the sampling error using independent trajectories, started from different initial conditions.[130, 131]

Although there are multiple ways to assess convergence, it can only be definitively shown that the sampling from MD is not converged. There is no way to prove convergence. A well-chosen measure of convergence can make a case that sampling is sufficient without proving it. On the other hand, a simulation without assessment of convergence has similar scientific value to a blot without loading controls: interpretable only at great peril. In practice, RNA has degrees of freedom that are quickly sampled, and other degrees of freedom that are sampled slowly (relative to simulation durations). The fast degrees of freedom can also be

coupled to the slow degrees of freedom, so apparent fast convergence can truly be slow as the molecule moves in its energy landscape.[128] Additionally, a simulation may appear converged based on estimates of statistical error, but MD samples a high dimensional space for RNA. Therefore there is always the possibility that additional sampling (by continuing the simulation) would cross an energy barrier and sample additional yet unobserved conformations, changing the statistical error estimates. In practice, it is important to use intuition about whether a simulation is long enough to address quantities of interest. A recent review explores these issues in greater detail, and also provides practical advice for considering convergence.[128]

4 Benchmarks and Applications

As presented above, there are many approximations in force fields and several methods for applying force fields. Thus, a variety of benchmarks are necessary for assessing the reliability of simulations. Several such benchmarks are presented below.

4.1 Equilibria

Important characteristics of RNA that can be predicted with MD simulations include the thermodynamics of folding and the structures present at equilibrium. The thermodynamics of base pairing and loop formation is usually coupled with experimental constraints and/or restraints for predicting RNA secondary structure.[7, 8, 9, 10, 11, 12] While robust experimental stabilities for Watson-Crick[132] and GU pairs [133] are available, there is limited data for the wide variety of loops. In principle, computations could predict loop stabilities. Additionally, RNAs form an ensemble of 3D structures. [16, 134, 135, 136, 137] Ultimately, computational methods should be able to accurately predict such ensembles.

4.2 Predicting Sequence Dependence of Folding Stability

Currently, a variety of methods and relatively simple systems are being used to benchmark computational predictions of folding stability, i.e. G . To see if computations with the previously popular AMBER FF99 (Figure 3) can reproduce experimental parameters for different stabilities of Watson-Crick pairs, the change in stability when one or two GC pairs are replaced with AU in the stem of a hairpin (Figure 6 e, f) was predicted. [138] The calculation was done with an umbrella sampling technique to calculate the predicted G at 300 K for going to a final state of completely extended single strand. The final state for experimental measurements is partially stacked single strands, so the differences between sequences were analyzed using thermodynamic cycles, assuming that there is no sequence dependence in relaxing the extended single strand to the partially stacked single strands. [139] Applicability of a sequence-independent worm like chain model[139] in compensating for different extents of single strand stacking was previously shown in an experimental study using single molecule pulling to measure G for unfolding five RNAs.[140] The predicted effect of one and two transitions from GC to AU in the hairpin agreed with nearest neighbor predictions[132] within 1.8 and 1.2 kcal/mol, respectively.[138]

A different type of comparison was made between experimental and predicted stabilities of tetramer duplexes with all GC or iGiC pairs [41] (Figures 5 and 7). In this case, the number

and type of hydrogen bonds are conserved but the partial charges on the bases change. [141] Thermodynamic integration was used to slowly transform the amino and carbonyl groups (Figure 5), which allows the final structure in the computation to be partially stacked single strands. New parametrization of α , β , γ , ϵ , ζ , and torsions[42, 110] for AMBER FF99 allowed agreement with experiment of between 1.3 and 2.6 kcal/mol at 300 K for the change in ΔG of duplex formation of four different duplexes when GC pairs were replaced with iGiC. For two sequences tested prior to reparametrization, the original and revised predicted ΔG disagreed with experiment by about 12 and 2 kcal/mol, respectively (Figure 7). Thus reparametrization improved predictions of relative equilibrium constants for duplex formation by roughly $\exp(10 \text{ kcal/mol/RT}) = \exp(104/600) = 107$ -fold, and the difference between experiment and prediction was reduced to roughly 30-fold. The force field revisions likely improved modeling of the unpaired single strands. Evidently, existing force fields and methods are close to being able to predict changes in thermodynamic stabilities of Watson-Crick paired helices due to chemical modification.

4.3 Predicting Structures and Ensembles for Loops

Many loops in RNA have function, such as binding ligands, forming tertiary interactions, and facilitating catalysis.[6, 142, 143, 144] Thus, computational methods that successfully predict structures and stabilities of loops will have major impact. Relatively small loops with GA pairs serve as an example of the current state of agreement with experiment. Such loops sometimes alternate between two structures, as revealed by NMR.[134] Usually, the GA pairs are largely in either a sheared or imino pairing (Figure 6). The computational challenge is to predict the thermodynamic stability, type, and extent of GA pairing. For example, Figure 8 shows four examples of NMR structures and thermodynamic stabilities for adjacent GA pairs flanked by GC or GU pairs in two possible orientations. The most stable motif, 5'GGAC/3'CAGG, is the only one with imino GA pairs. Replacing the C's with U's to give 5'GGAU/3'UAGG gives the least stable motif. That motif has sheared GA pairs and GU pairs with a single hydrogen bond (Figure 6B). The motifs, 5'CGAG/3'GAGC and 5'UGAG/3'GAGU have intermediate and similar thermodynamic stabilities, sheared GA pairs, and CG and UG pairs with three and two hydrogen bonds (Figure 6A), respectively. Evidently, both stacking and hydrogen bonding determine stability and structure, but neither is dominant. Thus, a force field requires subtle balance to predict the details of stability and structure for RNA loops. The change in GA pairing between imino in 5'GGAC and sheared in 5'GGAU and the single hydrogen bond GU pair in 5'GGAU were not predicted by AMBER FF10. [36]

A modification of the force field to allow a non-planar amino group for G in GA pairs, however, could predict the difference in GA hydrogen bonding patterns in 5'GGAC/3'CAGG and 5'GGAU/3'UAGG as well as the single hydrogen bond GU pair.[36] The modification was suggested by QM calculations on guanine[145] and previously improved comparisons of changes in GA pairing when GC flanking pairs are replaced with iGiC.[35, 146] Sequence patterns for terminal GA pairs in rRNA are also consistent with important interactions involving non-planar G amino groups, [147] although x-ray structures did not have sufficient resolution to confirm the non-planarity. The advantage of a QM approach is

illustrated by its application to iGiC base pairs and its revealing the non-planar potential of a G amino group.

UNCG, GNRA, and CUYG RNA tetraloops (N=any nucleotide; R=purine; Y=pyrimidine) provide another test of force fields. They are hyperabundant in ribosomal RNAs. [152] Solution and crystal structures have been obtained for all three classes. The stability of these tetraloops has been attributed to non-canonical base pairing and to hydrogen-bonding networks in the loops.

The first MD studies of tetraloops were on the GCAA [153] and UUCG tetraloops [154] and they continue to be challenging. With long-time scale simulations and enhanced sampling methods, tetraloop simulations reveal problems of force fields. Chen and Garcia [113] reported the folding of three tetraloops, including UUCG, by using a modified AMBER FF99 force field, with replica exchange molecular dynamics (REMD). They adjusted (Figure 3): (1) Leonard-Jones parameters as suggested by high-level quantum calculations[47] to correct excessive stacking propensities, (2) base-water interactions, and (3) the glycosidic torsion,. The resulting parameters could generate native base stacking and key noncanonical interactions in UUCG and GCAA, but not in CUUG.

The Cheatham group [115] used multidimensional replica exchange molecular dynamics (M-REMD) with a restrained base paired stem for simulations of the UUCG tetraloop. Two independent simulations sampled nearly identical conformational space, suggesting enough sampling to give convergence. Thus, some errors caused by incomplete sampling were removed. This improves force field assessment, and several AMBER force fields were tested, including FF12 (equivalent to FF10), FF12 + vdW modifications, [114] and FF99 + Chen-Garcia,[113] as well as the CHARMM forcefield.[118] Each force field generated an ensemble of structures, but structures very similar to that determined by NMR[151] were at most only 10% of an ensemble (Figure 9) using any of the force fields tested. Another study revealing limits of AMBER FF10 used a variant of metadynamics to predict folding free energies of the CCUUCGGG tetraloop.[155] The native structure of the tetraloop was predicted to have a vanishingly small population.

The systems described above are relatively restrained by hydrogen bonding and volume exclusion. Short, single stranded RNAs provide a different type of benchmark. [110] They lack hydrogen bonding between bases, are very flexible, and change conformations in less than a microsecond. [156, 157, 158, 159] Moreover, the limited number of atoms allows long simulations on current computers, thus allowing convergence. [115] The ensemble generated by simulations can be compared to NMR spectra. Such comparisons have been made for 5'GACC, [110, 160, 92, 161, 115, 108] 5'CAAU, [160, 108] 5'AAAA, [160, 108] 5'CCCC, [162, 92, 108] and 5'UUUU.[160, 108] Except for 5'UUUU, most simulations generate a large fraction of structures inconsistent with NMR spectra. For example, intercalated bases are often generated (Fig. 10). The Chen/Garcia version of FF99 generates a large fraction of conformations with a GC base pair for GACC.[115] Agreement with average distances and torsion angles determined from NMR spectra is less than 40% with FF10.[160] The results suggest that stacking and hydrogen bonding are not well balanced in current force fields.

While simulations of ensembles for small systems are problematic, encouraging results have been obtained with FF10 simulations of a kissing-loop complex containing 46 nucleotides of which all but six A's are in Watson-Crick pairs.[163] Simulations starting with a crystal structure having four of the A's bulged out of the interface helix[164] retain helical regions consistent with both x-ray and NMR structures,[165, 166, 167, 168] indicating that FF10 models Watson-Crick pairs well. Interestingly, the simulations irreversibly generate structures with all six A's inside the helix, which is consistent with NMR structures.[163] The results suggest that simulations with explicit water are best compared to solution NMR structures. The simulations, however, provide ensembles with expected values for chi torsions, which is inconsistent with the NMR structures. That comparison and others suggest that NMR structures can be improved by combining NMR data with simulations using current force fields. Such an approach has been able to improve two NMR structures of hairpins.[169]

As force fields improve, they can be benchmarked against many other structures. For example, structure prediction with a knowledge based force field and restraints from NMR chemical shifts was benchmarked against 23 noncanonical motifs.[170] In general, combining experimental restraints with current force field computations can provide reasonable 3D structures.

4.4 Predicting Rates of Conformational Changes

Current computational studies of RNA focus on predicting structures and thermodynamics. The kinetics of structural changes, however, are also important for folding and function. [171, 172, 173] Predicting rates of conformational changes requires long MD simulations. While many rates have been measured experimentally, they are usually too slow to serve as benchmarks for computations because of the limited time range currently accessible for simulations. Exceptions are the rates of stacking and unstacking of single stranded oligonucleotides, which are typically on the order of 10^7 s⁻¹. [156, 157, 158, 159] Methods for comparing RNA simulations to kinetic experiments are just starting to be developed. [174]

5 Future Perspectives

Force fields required for insights into RNA biology are less well developed than for protein or DNA. Proteins have less charge and fewer polar residues than RNA. Biologically important DNA structures are primarily double helices that store genetic information and are capped by quadruplexes. The 2'OH group in RNA allows folding into more complex structures that provide many different functions and can be targeted by therapeutics.[175, 176] Modified nucleotides are increasingly being discovered in RNAs [177] and structures are known for relatively few RNAs. Thus, there is much to discover about RNA structure and function. Reliably approximating the physics of interactions determining RNA structure can accelerate progress in understanding and controlling RNA. In particular, physical principles are basic and general and therefore applicable to completely new situations. For example, methods for generating a successful force field for natural RNA could be applied

to related synthetic polymers[178] being used for synthetic biology,[17, 18, 21, 22, 23, 179] for self-folding into desired structures, or as therapeutics.

One question is how accurate computational methods need to be in order to be useful to researchers. At human body temperature, 37 °C, equilibrium constants will be predicted within an order of magnitude if the ΔG° is predicted within 1.4 kcal/mol. Thus calculations within 1 kcal/mol will be useful. As described above, current methods can almost achieve this goal if the system is relatively rigid, such as a helix. Many critical parts of RNA, however, are in loops that are often flexible. In these cases, current methods need further improvement.

It is not surprising that it is difficult to develop a general force field for RNA.[180, 181, 182] The many types of interactions need to be adequately balanced, including stacking, various types of hydrogen bonds, and interaction with water. There is evidence that something about this balance is off in simulation studies of RNA.[160, 115] Electrostatic interactions are universal and long range. Moreover, they are variable depending on environment because water and the components of RNA, especially bases, are polarizable. Thus, partial charges will vary with environment. Polarizable force fields have been developed,[183, 184, 185] but are still at an early stage. Because double helices and loops present different environments and constraints, one can consider a somewhat different force field for balancing the interactions in each case.[35, 36] Methods like Hamiltonian replica exchange might permit dynamic force field transitions that allow the RNA to more accurately populate multiple basins in the same simulation, thereby allowing the simulator to more accurately interpret their relative free energies.

Much recent progress in predicting RNA properties has relied on improved parametrization of the energetics of backbone torsions as made possible by extensive QM calculations.[42, 43, 111, 186] Somewhat surprisingly, however, less effort has been expended on the conformational energetics of the ribose group. Crystal structures and NMR suggest the ribose is usually in either C3'-endo or C2'-endo conformations, but MD simulations often predict a wide range of conformations whose distributions do not match well with NMR or crystal structures. Because hydrogen bonds involving the 2'OH group are often found in 3D structures, improved parametrization of both the ribose ring and 2'OH may provide better predictions of RNA properties.[118]

Because RNA has a large variety of folds, validation of new force fields will require a large range of benchmarks, including prediction of thermodynamics and structure. There is a large database of thermodynamics for small loops as measured by optical melting.[187] The database of thermodynamics measured by single molecule pulling, however, is small.[140] The PDB database[15] of structures for small RNAs that can be used for benchmarks is rapidly increasing.[170] It is rare, however, to find cases where the thermodynamic and structural effects of small changes in sequence have been reported. Moreover, MD simulations typically predict ensembles while few experimental studies reveal details of conformational ensembles.[134, 135, 16, 136, 137] Ensembles are biologically important because interaction with other molecules can shift equilibria.

Because it is so easy to acquire sequence data, the ultimate application of computations is to predict all RNA properties from sequence. This would require a completely accurate understanding of the physics of RNA. A more realistic goal, however, is to combine computations with experiments to limit the detail required from experiments. For example, cryo-electron microscopy methods can almost provide structures close to those available by x-ray crystallography.[188, 189] Approximate force fields may be able to fill in the remaining gap, especially for crowded structures where conformational space is limited. In a similar way, experimental determinations of tertiary interactions coupled with MD can in principle provide detailed 3D structures.[190, 191, 192, 193] All atom molecular simulations have the potential to estimate an RNA's conformational ensemble, which is actually more important information for its biological function than a particular 3D structure; RNA is dynamic and its dynamism has much to do with its biology.

6 Conclusions

In principle, physics provides the knowledge to accurately predict RNA structure, stability, and dynamics from sequence. The increasing speed of computers permits more accurate QM calculations of energy as a function of structure for larger and larger components of RNA. [28] In turn, these results have allowed better parametrization of force fields that approximate QM with classical mechanics (Figure 3). The force fields are used in MD simulations to predict the ensemble of structures occupied by RNAs in solution. The MD data also permit prediction of the thermodynamics and dynamics of structural transitions. Advances in MD algorithms allow converged simulations of small RNAs so that the predicted structural ensembles and thermodynamics can be compared to experimental results. [115] New experimental benchmarks have been developed for such comparisons. The results of these comparisons show that revisions of parameters have improved predictions, but that there is much opportunity for further improvement. [115, 43, 41]

The ultimate goal of physics-based modeling of RNA is to accurately predict 3D structure and dynamics from sequence. This would accelerate understanding the roles of RNA in biology and the designing of therapeutics targeting RNA. Current computational methods can facilitate refinement and extension of experimental structures, and applications will expand as methods become more reliable. A major extension of force fields would be addition of polarizability to allow dynamic changes in partial charges during structural transitions. [183, 184] Advances in force fields could be benchmarked and guided by more experimental studies on model systems where a small change in covalent structure caused a large change in 3D structure and/or thermodynamics. Such studies could include effects of modified nucleotides, which are being rapidly identified in natural RNAs. [177] It will be important for experimentalists to report as much primary data as possible to allow MD simulations to be benchmarked against such data rather than against structural models.

6.1 Further Reading

Zuckerman, Daniel M. Statistical Physics of Biomolecules: An Introduction. CRC Press, 2010. Jensen, Frank. Introduction to Computational Chemistry, Second Edition. Wiley, 2006.

Acknowledgments

This work was funded by NIH grants R01 GM22939 (D.H.T.) and R01 GM076485 (D.H.M.). We thank Scott Kennedy for Figure 8 and Christina Bergonzo and Thomas E. Cheatham for providing Figure 9.

Conflict of interest: The authors have declared no conflicts of interest for this article.

References

1. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. <http://dx.doi.org/10.1038/nature05874>. [PubMed: 17571346]
2. Consortium R. RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Research*. 2014; gku991.
3. Michel F, Jacquier A, Dujon B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie*. 1982; 64:867–881. [PubMed: 6817818]
4. Waring RB, Davies RW. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing—a review. *Gene*. 1984; 28:277–291. [PubMed: 6086458]
5. Cech TR, Damberger SH, Gutell RR. Representation of the secondary and tertiary structure of group I introns. *Nature Structural & Molecular Biology*. 1994; 1:273–280.
6. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science*. 1996; 273:1678. <http://science.sciencemag.org/content/273/5282/1678.abstract>. [PubMed: 8781224]
7. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Research*. 2008; 36:W70–W74. [PubMed: 18424795]
8. Chen JL, Bellaousov S, Tubbs JD, Kennedy SD, Lopez MJ, Mathews DH, Turner DH. Nuclear Magnetic Resonance-Assisted Prediction of Secondary Structure for RNA: Incorporation of Direction-Dependent Chemical Shift Constraints. *Biochemistry*. 2015; 54:6769–6782. [PubMed: 26451676]
9. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:7287–7292. [PubMed: 15123812]
10. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:97–102. [PubMed: 19109441]
11. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:5498–5503. [PubMed: 23503844]
12. Wu Y, Shi B, Ding X, Liu T, Hu X, Yip KY, Yang ZR, Mathews DH, Lu ZJ. Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Research*. 2015; gkv706.
13. Mathews DH, Turner DH. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *Journal of molecular biology*. 2002; 317:191–203. [PubMed: 11902836]
14. Fu Y, Sharma G, Mathews DH. Dynalign II: common secondary structure prediction for RNA homologs with domain insertions. *Nucleic acids research*. 2014; 42:13939–13948. [PubMed: 25416799]
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Research*. 2000; 28:235–242. [PubMed: 10592235]
16. Salmon L, Yang S, Al-Hashimi HM. Advances in the Determination of Nucleic Acid Conformational Ensembles. *Annual Review of Physical Chemistry*. 2014; 65:293–316. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4748180/>.

17. Xiao H, Schultz PG. At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harbor perspectives in biology*. 2016; 8:a023945. [PubMed: 27413101]
18. Xiao H, Nasertorabi F, Choi S-h, Han GW, Reed SA, Stevens RC, Schultz PG. Exploring the potential impact of an expanded genetic code on protein function. *Proceedings of the National Academy of Sciences*. 2015; 112:6961–6966.
19. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, SantaLucia J. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *Journal of Chemical Theory and Computation*. 2007; 3:1464–1475. [PubMed: 26633217]
20. Xu Y, Vanommeslaeghe K, Aleksandrov A, MacKerell AD, Nilsson L. Additive CHARMM force field for naturally occurring modified ribonucleotides. *Journal of Computational Chemistry*. Jan. 2016 37:896–912. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4801715/>. [PubMed: 26841080]
21. Winnacker M, Kool ET. Artificial Genetic Sets Composed of SizeExpanded Base Pairs. *Angewandte Chemie International Edition*. 2013; 52:12498–12508. [PubMed: 24249550]
22. Karalkar NB, Leal NA, Kim M-S, Bradley KM, Benner SA. Synthesis and Enzymology of 2-Deoxy-7-deazaisoguanosine Triphosphate and Its Complement: A Second Generation Pair in an Artificially Expanded Genetic Information System. *ACS synthetic biology*. 2016
23. Chen T, Hongdilkul N, Liu Z, Thirunavukarasu D, Romesberg FE. The expanding world of DNA and RNA. *Current Opinion in Chemical Biology*. 2016; 34:80–87. [PubMed: 27565457]
24. Kührová P, Otyepka M, Šponer J, Banáš P. Are Waters around RNA More than Just a Solvent? – An Insight from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*. Jan.2014 10:401–411. <http://dx.doi.org/10.1021/ct400663s>. [PubMed: 26579919]
25. Hill JG, Platts JA. Calculating stacking interactions in nucleic acid base-pair steps using spin-component scaling and local second order Moller-Plesset perturbation theory. *Physical Chemistry Chemical Physics*. 2008; 10:2785–2791. <http://dx.doi.org/10.1039/B718691F>. [PubMed: 18464995]
26. Morgado CA, Svozil D, Turner DH, Šponer J. Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes. *Physical Chemistry Chemical Physics*. Sep.2012 14:12580–12591. <http://dx.doi.org/10.1039/c2cp40556c>. [PubMed: 22722325]
27. Šponer J, Jurečka P, Marchan I, Luque FJ, Orozco M, Hobza P. Nature of Base Stacking: Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps. *Chemistry – A European Journal*. 2006; 12:2854–2865. <http://dx.doi.org/10.1002/chem.200501239>.
28. Kruse H, Mladek A, Gkionis K, Hansen A, Grimme S, Sponer J. Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *Journal of Chemical Theory and Computation*. Oct.2015 11:4972–4991. <http://dx.doi.org/10.1021/acs.jctc.5b00515>. [PubMed: 26574283]
29. Stone, A. *The Theory of Intermolecular Forces*. OUP; Oxford: 2013. *The Theory of Intermolecular Forces*. <https://books.google.com/books?id=yOHe5Lpj71MC>
30. Jensen, F. *Introduction to Computational Chemistry*. 2. Wiley; 2006.
31. Lifson S, Warshel A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *The Journal of Chemical Physics*. 1968; 49:5116. <http://dx.doi.org/10.1063/1.1670007>.
32. Ponder, JW., Case, D. *Protein Simulations*, volume 66 of *Advances in Protein Chemistry*. Academic Press; 2003. *Force Fields for Protein Simulations*; p. 27-85.<http://www.sciencedirect.com/science/article/pii/S006532330366002X>
33. Case, D., Betz, R., Botello-Smith, W., Cerutti, D., Cheatham, TE., III, Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, KM., Monard, G., Nguyen, H., Nguyen, H., Omelyan, I., Onufriev, A., Roe, DR., Roitberg, AE., Sagui, C., Simmerling, C., Swails, J., Walker, R., Wang, J., Wolf, R., Wu, X., Xiao, L., York, D., Kollman, P. *AMBER 2016*. University of California; San Francisco: 2016.
34. Sponer J, Leszczynski J, Hobza P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers*. 2001; 61:3–31. <http://dx.doi.org/3.0.CO;2-4>.

35. Yildirim I, Stern HA, Sponer J, Spackova N, Turner DH. Effects of Restrained Sampling Space and Nonplanar Amino Groups on Free-Energy Predictions for RNA with Imino and Sheared Tandem GA Base Pairs Flanked by GC, CG, iGiC or iCiG Base Pairs. *Journal of Chemical Theory and Computation*. Aug.2009 5:2088–2100. <http://dx.doi.org/10.1021/ct800540c>. [PubMed: 20090924]
36. Aytenfisu AH, Spasic A, Seetin MG, Serafini J, Mathews DH. Modified AMBER force field correctly models the conformational preference for tandem GA pairs in RNA. *Journal of Chemical Theory and Computation*. 2014; 10:1292–1301. [PubMed: 24803859]
37. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*. 1995; 117:5179–5197. <http://dx.doi.org/10.1021/ja00124a002>.
38. MacKerell ADJ, Bashford D, Bellott M, Dunbrack RLJ, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*. 1998; 102:3586–3616. <http://dx.doi.org/10.1021/jp973084f>. [PubMed: 24889800]
39. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, HersHKovits E, Williams LD, Keating KS, Pyle AM, Micallef D, Westbrook J, Berman HM. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*. 2008; 14:465–481. <http://rnajournal.cshlp.org/content/14/3/465.abstract>. [PubMed: 18192612]
40. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*. 2000; 21:1049–1074. [http://dx.doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](http://dx.doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F).
41. Yildirim I, Kennedy SD, Stern HA, Hart JM, Kierzek R, Turner DH. Revision of AMBER torsional parameters for RNA improves free energy predictions for tetramer duplexes with GC and iGiC base pairs. *Journal of Chemical Theory and Computation*. 2011; 8:172–181. [PubMed: 22249447]
42. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE III, Lughton CA, Orozco M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Conformers. *Biophysical Journal*. 2007; 92:3817–3829. <http://www.sciencedirect.com/science/article/pii/S0006349507711827>. [PubMed: 17351000]
43. Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Thomas E, Cheatham I, Jurečka P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation*. 2011; 7:2886–2902. <http://dx.doi.org/10.1021/ct200162x>. [PubMed: 21921995]
44. Ts'o, PO. Bases, Nucleosides and Nucleotides. In: Ts'o, P., editor. *Basic Principles in Nucleic Acid Chemistry*. Vol. I. Academic Press; 1974. p. 453–584.
45. Tribolet R, Sigel H. Self-association and protonation of adenosine 5'-monophosphate in comparison with its 2'- and 3'-analogues and tubercidin 5'-monophosphate (7-deaza-AMP). *European Journal of Biochemistry*. 1987; 163:353–363. <http://dx.doi.org/10.1111/j.1432-1033.1987.tb10807.x>. [PubMed: 3028802]
46. Solie T, Schellman J. The interaction of nucleosides in aqueous solution. *Journal of Molecular Biology*. 1968; 33:61–77. <http://www.sciencedirect.com/science/article/pii/0022283668902817>. [PubMed: 5646654]
47. Sponer J, Sponer JE, Mládek A, Jurečka P, Banáš P, Otyepka M. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*. Dec.2013 99:978–988. <http://dx.doi.org/10.1002/bip.22322>. [PubMed: 23784745]
48. Grimme S. Do Special Noncovalent – Stacking Interactions Really Exist? *Angewandte Chemie International Edition*. 2008; 47:3430–3434. <http://dx.doi.org/10.1002/anie.200705157>. [PubMed: 18350534]
49. Burkard ME, Kierzek R, Turner DH. Thermodynamics of unpaired terminal nucleotides on short {RNA} helices correlates with stacking at helix termini in larger {RNAs1}. *Journal of Molecular*

- Biology. 1999; 290:967–982. www.sciencedirect.com/science/article/pii/S0022283699929066. [PubMed: 10438596]
50. Wang Q, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal J-P, Ren P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *Journal of Chemical Theory and Computation*. 2015; 11:2609–2618. <http://dx.doi.org/10.1021/acs.jctc.5b00267>. [PubMed: 26413036]
51. Parker TM, Sherrill CD. Assessment of Empirical Models versus High-Accuracy Ab Initio Methods for Nucleobase Stacking: Evaluating the Importance of Charge Penetration. *Journal of Chemical Theory and Computation*. 2015; 11:4197–4204. <http://dx.doi.org/10.1021/acs.jctc.5b00588>. [PubMed: 26575915]
52. Lee C, McCammon JA, Rossky PJ. The structure of liquid water at an extended hydrophobic surface. *The Journal of Chemical Physics*. 1984; 80:4448–4455. <http://scitation.aip.org/content/aip/journal/jcp/80/9/10.1063/1.447226>.
53. Lee SH, Rossky PJ. A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces - a molecular dynamics simulation study. *The Journal of Chemical Physics*. 1994; 100:3334–3345. <http://scitation.aip.org/content/aip/journal/jcp/100/4/10.1063/1.466425>.
54. Tavagnacco L, Di Fonzo S, D'Amico F, Masciovecchio C, Brady JW, Cesaro A. Stacking of purines in water: the role of dipolar interactions in caffeine. *Physical Chemistry Chemical Physics*. 2016; 18:13478–13486. <http://dx.doi.org/10.1039/C5CP07326J>. [PubMed: 27127808]
55. Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*. Sep.2005 437:640–647. <http://dx.doi.org/10.1038/nature04162>. [PubMed: 16193038]
56. Zangi R, Berne BJ. Temperature Dependence of Dimerization and Dewetting of Large-Scale Hydrophobes: A Molecular Dynamics Study. *The Journal of Physical Chemistry B*. 2008; 112:8634–8644. <http://dx.doi.org/10.1021/jp802135c>. [PubMed: 18582012]
57. Lennard-Jones JE. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 1924; 106:463–477. <http://rspa.royalsocietypublishing.org/content/106/738/463>.
58. Hart JR, Rappé AK. van der Waals functional forms for molecular simulations. *The Journal of Chemical Physics*. 1992; 97:1109–1115. <http://scitation.aip.org/content/aip/journal/jcp/97/2/10.1063/1.463290>.
59. Berthelot D. Sur le mélange des gaz. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*. 1898; 126:1703–1855. <http://visualiseur.bnf.fr/Document/CadresPage?O=NUMM-3082I=1703>.
60. Lorentz HA. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der Physik*. 1881; 248:127–136. <http://dx.doi.org/10.1002/andp.18812480110>.
61. Waldman M, Hagler A. New combining rules for rare gas van der Waals parameters. *Journal of Computational Chemistry*. 1993; 14:1077–1084. <http://dx.doi.org/10.1002/jcc.540140909>.
62. Delhommelle J, Millié P. Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation. *Molecular Physics*. 2001; 99:619–625. <http://dx.doi.org/10.1080/00268970010020041>.
63. Jorgensen W, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*. 1988; 110:1657–1666. <http://dx.doi.org/10.1021/ja00214a001>. [PubMed: 27557051]
64. Jorgensen W, Maxwell D, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*. 1996; 118:11225–11236.
65. Foloppe N, MacKerell AD Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*. 2000; 21:86–104. [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<86::AID-JCC2>3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G).

66. Momany FA. Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid. *The Journal of Physical Chemistry*. 1978; 82:592–601. <http://dx.doi.org/10.1021/j100494a019>.
67. Cox SR, Williams DE. Representation of the molecular electrostatic potential by a net atomic charge model. *Journal of Computational Chemistry*. 1981; 2:304–323. <http://dx.doi.org/10.1002/jcc.540020312>.
68. Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*. 1984; 5:129–145. <http://dx.doi.org/10.1002/jcc.540050204>.
69. Faerman CH, Price SL. A transferable distributed multipole model for the electrostatic interactions of peptides and amides. *Journal of the American Chemical Society*. 1990; 112:4915–4926. <http://dx.doi.org/10.1021/ja00168a043>.
70. Price SL, Faerman CH, Murray CW. Toward accurate transferable electrostatic models for polypeptides: A distributed multipole study of blocked amino acid residue charge distributions. *Journal of Computational Chemistry*. 1991; 12:1187–1197. <http://dx.doi.org/10.1002/jcc.540121005>.
71. Reynolds CA, Essex JW, Richards WG. Atomic charges for variable molecular conformations. *Journal of the American Chemical Society*. 1992; 114:9075–9079. <http://dx.doi.org/10.1021/ja00049a045>.
72. Cornell WD, Cieplak P, Bayly CI, Kollman PA. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society*. 1993; 115:9620–9631. <http://dx.doi.org/10.1021/ja00074a030>.
73. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*. 1993; 97:10269–10280. <http://dx.doi.org/10.1021/j100142a004>.
74. Szabo, A., Ostlund, N. *Dover Books on Chemistry*. Dover Publications; 1989. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. <https://books.google.com/books?id=6mV9gYzEkgIC>
75. Cieplak P, Cornell WD, Bayly C, Kollman PA. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry*. 1995; 16:1357–1377. <http://dx.doi.org/10.1002/jcc.540161106>.
76. Besler BH, Merz KM, Kollman PA. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*. 1990; 11:431–439. <http://dx.doi.org/10.1002/jcc.540110404>.
77. Arunan E, Desiraju GR, Klein RA, Sadlej J, Sheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, Kjaergaard HG, Legon AC, Mennucci B, Nesbitt DJ. Definition of the hydrogen bond. *Pure and Applied Chemistry*. Oct.2011 83:1637–1641. <http://www.iupac.org/publications/pac/pdf/2011/pdf/8308x1637.pdf>.
78. Gilli P, Bertolasi V, Ferretti V, Gilli G. Evidence for resonance-assisted hydrogen bonding. 4. Covalent nature of the strong homonuclear hydrogen bond. Study of the O–H–O system by crystal structure correlation methods. *Journal of the American Chemical Society*. 1994; 116:909–915. <http://dx.doi.org/10.1021/ja00082a011>.
79. Gilli P, Pretto L, Bertolasi V, Gilli G. Predicting Hydrogen-Bond Strengths from Acid–base Molecular Properties. The pKa Slide Rule: Toward the Solution of a Long-Lasting Problem. *Accounts of Chemical Research*. 2009; 42:33–44. <http://dx.doi.org/10.1021/ar800001k>. [PubMed: 18921985]
80. Guillaumes L, Simon S. Dealing with Quasi-Ring Formation by Two Hydrogen Bonds. Cooperativity Analysis with Delocalization Indices. *The Journal of Physical Chemistry A*. 2014; 118:9727–9733. <http://dx.doi.org/10.1021/jp507133n>. [PubMed: 25229157]
81. Freier SM, Sugimoto N, Sinclair A, Alkema D, Neilson T, Kierzek R, Caruthers MH, Turner DH. Stability of XGCGCp, GCGCYp, and XGCGCYp helices: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry*. 1986; 25:3214–3219. [PubMed: 3730357]
82. Turner DH, Sugimoto N, Kierzek R, Dreiker SD. Free energy increments for hydrogen bonds in nucleic acid base pairs. *Journal of the American Chemical Society*. 1987; 109:3783–3785. <http://dx.doi.org/10.1021/ja00246a047>.

83. SantaLucia J, Kierzek R, Turner D. Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*. 1992; 256:217–219. <http://science.sciencemag.org/content/256/5054/217>. [PubMed: 1373521]
84. Auffinger P, Westhof E. Simulations of the molecular dynamics of nucleic acids. *Current Opinion in Structural Biology*. 1998; 8:227–236. <http://www.sciencedirect.com/science/article/pii/S0959440X98800444>. [PubMed: 9631298]
85. Egli M, Portmann S, Usman N. RNA Hydration: A Detailed Look. *Biochemistry*. 1996; 35:8489–8494. <http://dx.doi.org/10.1021/bi9607214>. [PubMed: 8679609]
86. Auffinger P, Westhof E. Rules governing the orientation of the 2'-hydroxyl group in RNA1. *Journal of Molecular Biology*. 1997; 274:54–63. <http://www.sciencedirect.com/science/article/pii/S002228369791370X>. [PubMed: 9398515]
87. Yoon J, Lin J-C, Hyeon C, Thirumalai D. Dynamical Transition and Heterogeneous Hydration Dynamics in RNA. *The Journal of Physical Chemistry B*. 2014; 118:7910–7919. <http://dx.doi.org/10.1021/jp500643u>. [PubMed: 24762118]
88. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983; 79:926–935.
89. Berendsen, HJC., Postma, JPM., van Gunsteren, WF., Hermans, J. *Interaction Models for Water in Relation to Protein Hydration*. Springer Netherlands; Dordrecht: 1981. p. 331-342. http://dx.doi.org/10.1007/978-94-015-7658-1_21
90. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *The Journal of Physical Chemistry*. 1987; 91:6269–6271. <http://dx.doi.org/10.1021/j100308a038>.
91. Izadi S, Anandakrishnan R, Onufriev AV. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters*. 2014; 5:3863–3871. <http://dx.doi.org/10.1021/jz501780a>. [PubMed: 25400877]
92. Bergonzo C, Cheatham TEI. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *Journal of Chemical Theory and Computation*. 2015; 11:3969–3972. <http://dx.doi.org/10.1021/acs.jctc.5b00444>. [PubMed: 26575892]
93. Sklenovský P, Florová P, Banáš P, Réblová K, Lankaš F, Otyepka M, Šponer J. Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *Journal of Chemical Theory and Computation*. Sep.2011 7:2963–2980. <http://dx.doi.org/10.1021/ct200204t>. [PubMed: 26605485]
94. Draper DE. A guide to ions and RNA structure. *RNA*. 2004; 10:335–343. <http://rnajournal.cshlp.org/content/10/3/335.abstract>. [PubMed: 14970378]
95. Beveridge DL, Barreiro G, Byun KS, Case DA III, TEC, Dixit SB, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Seibert E, Sklenar H, Stoll G, Thayer KM, Varnai P, Young MA. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of {DNA} Oligonucleotides. I. Research Design and Results on d(CpG) Steps. *Biophysical Journal*. 2004; 87:3799– 3813. <http://www.sciencedirect.com/science/article/pii/S0006349504738473>. [PubMed: 15326025]
96. Dixit SB, Beveridge DL, Case DA 3rd, TEC, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Sklenar H, Thayer KM, Varnai P. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of {DNA} Oligonucleotides. II: Sequence Context Effects on the Dynamical Structures of the 10 Unique Dinucleotide Steps. *Biophysical Journal*. 2005; 89:3721–3740. <http://www.sciencedirect.com/science/article/pii/S0006349505730186>. [PubMed: 16169978]
97. Lipfert J, Doniach S, Das R, Herschlag D. Understanding nucleic acid–ion interactions. *Annual Review of Biochemistry*. 2014; 83:813–841.
98. Misra VK, Draper DE. On the role of magnesium ions in RNA stability. *Biopolymers*. 1998; 48:113–135. [http://dx.doi.org/10.1002/\(SICI\)1097-0282\(1998\)48:2<113::AID-BIP3>3.0.CO;2-Y](http://dx.doi.org/10.1002/(SICI)1097-0282(1998)48:2<113::AID-BIP3>3.0.CO;2-Y). [PubMed: 10333741]
99. Grossfield AM, Ren P, Ponder JW. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society*. 2003; 125:15671–15682. <http://dx.doi.org/10.1021/ja037005r>. [PubMed: 14664617]

100. Piquemal J-P, Perera L, Cisneros GA, Ren P, Pedersen LG, Darden TA. Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *The Journal of Chemical Physics*. 2006; 125 <http://scitation.aip.org/content/aip/journal/jcp/125/5/10.1063/1.2234774>.
101. Kowerko D, König SLB, Skilandat M, Kruschel D, Hadzic MCAS, Cardo L, Sigel RKO. Cation-induced kinetic heterogeneity of the intron–exon recognition in single group II introns. *Proceedings of the National Academy of Sciences*. 2015; 112:3403–3408. <http://www.pnas.org/content/112/11/3403.abstract>.
102. Bergonzo C., Hall, KB., Cheatham, TE. Divalent Ion Dependent Conformational Changes in an RNA Stem-Loop Observed by Molecular Dynamics. *Journal of Chemical Theory and Computation*. 2016. 0:null<http://dx.doi.org/10.1021/acs.jctc.6b00173>
103. Casalino L, Palermo G, Abdurakhmonova N, Rothlisberger U, Magistrato A. Development of Site-Specific Mg²⁺–RNA Force Field Parameters: A Dream or Reality? Guidelines from Combined Molecular Dynamics and Quantum Mechanics Simulations. *Journal of Chemical Theory and Computation*. 2017; 13:340–352. <http://dx.doi.org/10.1021/acs.jctc.6b00905>. [PubMed: 28001405]
104. Anandakrishnan R, Baker C, Izadi S, Onufriev AV. Point Charges Optimally Placed to Represent the Multipole Expansion of Charge Distributions. *PLoS ONE*. 2013; 8:1–13. <http://dx.doi.org/10.1371>.
105. White AD, Dama JF, Voth GA. Designing Free Energy Surfaces That Match Experimental Data with Metadynamics. *Journal of Chemical Theory and Computation*. 2015; 11:2451–2460. <http://dx.doi.org/10.1021/acs.jctc.5b00178>. [PubMed: 26575545]
106. Marinelli F, Faraldo-Gómez JD. Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophysical Journal*. 2015; 108:2779–2782. <http://www.sciencedirect.com/science/article/pii/S0006349515005366>. [PubMed: 26083917]
107. Gil-Ley A, Bottaro S, Bussi G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *Journal of Chemical Theory and Computation*. 2016; 12:2790–2798. <http://dx.doi.org/10.1021/acs.jctc.6b00299>. [PubMed: 27153317]
108. Aytenfisu, AH., Spasic, A., Grossfield, A., Stern, HA., Mathews, DH. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. *Journal of Chemical Theory and Computation*. 2017. 0:null<http://dx.doi.org/10.1021/acs.jctc.6b00870>
109. Laio A, Parrinello M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*. Jul.2002 99:12562–12566. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC130499/>. [PubMed: 12271136]
110. Yildirim I, Stern HA, Tubbs JD, Kennedy SD, Turner DH. Benchmarking AMBER Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised chi Torsions. *The Journal of Physical Chemistry B*. Jul.2011 115:9261–9270. <http://dx.doi.org/10.1021/jp2016006>. [PubMed: 21721539]
111. Yildirim I, Stern HA, Kennedy SD, Tubbs JD, Turner DH. Reparameterization of RNA chi torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of Chemical Theory and Computation*. 2010; 6:1520–1531. [PubMed: 20463845]
112. Yildirim I, Kennedy SD, Stern HA, Hart JM, Kierzek R, Turner DH. Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iGiC Base Pairs. *Journal of Chemical Theory and Computation*. 2012; 8:172–181. <http://dx.doi.org/10.1021/ct200557r>. [PubMed: 22249447]
113. Chen AA, Garca AE. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proceedings of the National Academy of Sciences*. 2013; 110:16820–16825.
114. Steinbrecher T, Latzer J, Case D. Revised AMBER parameters for bioorganic phosphates. *Journal of Chemical Theory and Computation*. 2012; 8:4405–4412. [PubMed: 23264757]
115. Bergonzo C, Henriksen NM, Roe DR, Cheatham TE. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*. 2015; 21:1578–1590. <http://rnajournal.cshlp.org/content/21/9/1578.abstract>. [PubMed: 26124199]

116. MacKerell AD, Wiorkiewicz-Kuczera J, Karplus M. An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical Society*. 1995; 117:11946–11975. <http://dx.doi.org/10.1021/ja00153a017>.
117. MacKerell AD, Banavali NK. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry*. 2000; 21:105–120. [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<105::AID-JCC3>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P).
118. Denning EJ, Priyakumar UD, Nilsson L, Mackerell AD. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of Computational Chemistry*. 2011; 32:1929–1943. <http://dx.doi.org/10.1002/jcc.21777>. [PubMed: 21469161]
119. Cramer, C. *Essentials of Computational Chemistry: Theories and Models*. Wiley; 2005. <https://books.google.com/books?id=BKJcbvzI0RsC>
120. Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*. Jul.1967 159:98–103. <http://link.aps.org/doi/10.1103/PhysRev.159.98>.
121. Becker, O., MacKerell, A., Roux, B., Watanabe, M. *Computational Biochemistry and Biophysics*. Taylor & Francis; 2001. <https://books.google.com/books?id=wQt5mgD4x9YC>
122. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*. 1977; 23:327– 341. <http://www.sciencedirect.com/science/article/pii/0021999177900985>.
123. Andersen HC. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*. 1983; 52:24– 34. <http://www.sciencedirect.com/science/article/pii/0021999183900141>.
124. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*. Oct.1992 13:1011–1021. <http://dx.doi.org/10.1002/jcc.540130812>.
125. Grossfield, AM. WHAM: The Weighted Histogram Analysis Method, 2.0.9 edition. 2016. <http://membrane.urmc.rochester.edu/content/wham>
126. Simonson T, Carlsson J, Case DA. Proton Binding to Proteins: pKa Calculations with Explicit and Implicit Solvent Models. *Journal of the American Chemical Society*. 2004; 126:4167–4180. <http://dx.doi.org/10.1021/ja039788m>. [PubMed: 15053606]
127. Press, W., Teukolsky, S., Vetterling, W., Flannery, B. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press; 1992. Number bk. 4 in *Numerical recipes in C: the Art of Scientific Computing* <https://books.google.com/books?id=4t-sybVuoqoC>
128. Grossfield A, Zuckerman DM. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annual Reports in Computational Chemistry*. 2009; 5:23–48. [PubMed: 20454547]
129. Flyvbjerg H, Petersen HG. Error estimates on averages of correlated data. *The Journal of Chemical Physics*. 1989; 91:461–466.
130. Zhang X, Walker RC, Phizicky EM, Mathews DH. Influence of Sequence and Covalent Modifications on Yeast tRNA Dynamics. *Journal of Chemical Theory and Computation*. 2014; 10:3473–3483. [PubMed: 25136272]
131. Auffinger P, Louise-May S, Westhof E. Multiple molecular dynamics simulations of the anticodon loop of tRNA^{Asp} in aqueous solution with counterions. *Journal of the American Chemical Society*. 1995; 117:6720–6726.
132. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs†. *Biochemistry*. 1998; 37:14719–14735. (Print) 0006–2960. [PubMed: 9778347]
133. Chen JL, Dishler AL, Kennedy SD, Yildirim I, Liu B, Turner DH, Serra MJ. Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*. 2012; 51:3508–3522. [PubMed: 22490167]

134. Hammond NB, Tolbert BS, Kierzek R, Turner DH, Kennedy SD. RNA Internal Loops with Tandem AG Pairs: The Structure of the 5'G AG U/3'U GA G Loop Can Be Dramatically Different from Others, Including 5'A AG U/3'U GA A. *Biochemistry*. 2010; 49:5817–5827. [PubMed: 20481618]
135. Kennedy SD, Kierzek R, Turner DH. Novel conformation of an RNA structural switch. *Biochemistry*. 2012; 51:9257–9259. [PubMed: 23134175]
136. Lee J, Dethoff EA, Al-Hashimi HM. Invisible RNA state dynamically couples distant motifs. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:9485–9490. [PubMed: 24979799]
137. Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature*. 2015; 519:315–320. <http://dx.doi.org/10.1038/nature14227>. [PubMed: 25762137]
138. Spasic A, Serafini J, Mathews DH. The Amber ff99 Force Field Predicts Relative Free Energy Changes for RNA Helix Formation. *Journal of Chemical Theory and Computation*. 2012; 8:2497–2505. <http://dx.doi.org/10.1021/ct300240k>. [PubMed: 23112748]
139. Flory PJ, Volkenstein M. Statistical mechanics of chain molecules. *Biopolymers*. 1969; 8:699–700. <http://dx.doi.org/10.1002/bip.1969.360080514>.
140. Viereggs J, Cheng W, Bustamante C, Tinoco I. Measurement of the effect of monovalent cations on RNA hairpin stability. *Journal of the American Chemical Society*. 2007; 129:14966–14973. [PubMed: 17997555]
141. Yildirim I, Turner DH. RNA challenges for computational chemists. *Biochemistry*. 2005; 44:13225–13234. [PubMed: 16201748]
142. Wyatt, JR., Tinoco, I. RNA structural elements and RNA function. In: Gesteland, RF., Atkins, JF., editors. *The RNA World*. Vol. chapter 18. Cold Spring Harbor Laboratory Press; 1993. p. 465–496.
143. Atkins, JF., Gesteland, RF., Cech, T. *RNA worlds: from life's origins to diversity in gene regulation*. Cold Spring Harbor Laboratory Press; New York: 2011.
144. Bevilacqua PC, Bloise JM. Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu Rev Phys Chem*. 2008; 59:79–103. [PubMed: 17937599]
145. Hobza P, Šponer J. Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chemical Reviews*. 1999; 99:3247–3276. [PubMed: 11749516]
146. Chen G, Kierzek R, Yildirim I, Krugh TR, Turner DH, Kennedy SD. Stacking effects on local structure in RNA: changes in the structure of tandem GA pairs when flanking GC pairs are replaced by isoG-isoC pairs. *The Journal of Physical Chemistry B*. 2007; 111:6718–6727. [PubMed: 17411085]
147. Šponer J, Mokdad A, Šponer JE, Špačková N, Leszczynski J, Leontis NB. Unique tertiary and neighbor interactions determine conservation patterns of Cis Watson–Crick A/G base-pairs. *Journal of Molecular Biology*. 2003; 330:967–978. [PubMed: 12860120]
148. Tolbert BS, Kennedy SD, Schroeder SJ, Krugh TR, Turner DH. NMR Structures of (rGC U GA G GCU) 2 and (rGC G GA U GCU) 2: Probing the Structural Features That Shape the Thermodynamic Stability of GA Pairs. *Biochemistry*. 2007; 46:1511–1522. [PubMed: 17279616]
149. Wu M, Turner DH. Solution structure of (rGCGGACGC) 2 by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry*. 1996; 35:9677–9689. [PubMed: 8703939]
150. Santa Lucia J Jr, Turner DH. Structure of (rGGCGAGCC) 2 in solution from NMR and restrained molecular dynamics. *Biochemistry*. 1993; 32:12612–12623. [PubMed: 8251479]
151. Nozinovic S, Fürtig B, Jonker HRA, Richter C, Schwalbe H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Research*. 2010; 38:683–694. <http://nar.oxfordjournals.org/content/38/2/683.abstract>. [PubMed: 19906714]
152. Woese CR, Winker S, Gutell RR. Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proceedings of the National Academy of Sciences*. 1990; 87:8467–8471. <http://www.pnas.org/content/87/21/8467.abstract>.

153. Zichi DA. Molecular Dynamics of RNA with the OPLS Force Field. Aqueous Simulation of a Hairpin Containing a Tetranucleotide Loop. *Journal of the American Chemical Society*. 1995; 117:2957–2969. <http://dx.doi.org/10.1021/ja00116a001>.
154. Miller JL, Kollman PA. Theoretical studies of an exceptionally stable {RNA} tetraloop: observation of convergence from an incorrect {NMR} structure to the correct one using unrestrained molecular dynamics I. *Journal of Molecular Biology*. 1997; 270:436–450. <http://www.sciencedirect.com/science/article/pii/S002228369791113X>. [PubMed: 9237909]
155. Bottaro S, Banáš P, Šponer J, Bussi G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *The Journal of Physical Chemistry Letters*. 2016; 7:4032–4038. <http://dx.doi.org/10.1021/acs.jpclett.6b01905>. [PubMed: 27661094]
156. Pöschke D. The nature of stacking interactions in polynucleotides. Molecular states in oligo- and polyribocytidylic acids by relaxation analysis. *Biochemistry*. 1976; 15:1495–1499. [PubMed: 1259951]
157. Pöschke D. Molecular states in single-stranded adenylate chains by relaxation analysis. *Biopolymers*. 1978; 17:315–323.
158. Dewey T, Turner DH. Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry*. 1979; 18:5757–5762. [PubMed: 518868]
159. Freier SM, Hill K, Dewey T, Marky LA, Breslauer KJ, Turner DH. Solvent effects on the kinetics and thermodynamics of stacking in poly (cytidylic acid). *Biochemistry*. 1981; 20:1419–1426. [PubMed: 7225340]
160. Condon DE, Kennedy SD, Mort BC, Kierzek R, Yildirim I, Turner DH. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *Journal of Chemical Theory and Computation*. 2015; 11:2729–2742. <http://dx.doi.org/10.1021/ct501025q>. [PubMed: 26082675]
161. Bergonzo C, Henriksen NM, Roe DR, Swails JM, Roitberg AE, Cheatham TE. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *Journal of Chemical Theory and Computation*. 2014; 10:492–499. (Print) 1549–9618 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3893832/pdf/ct400862k.pdf>. [PubMed: 24453949]
162. Tubbs JD, Condon DE, Kennedy SD, Hauser M, Bevilacqua PC, Turner DH. The nuclear magnetic resonance of CCCC RNA reveals a right-handed helix, and revised parameters for AMBER force field torsions improve structural predictions from molecular dynamics. *Biochemistry*. 2013; 52:996–1010. [PubMed: 23286901]
163. Havrila M, Zgarbová M, Jurečka P, Banáš P, Krepl M, Otyepka M, Šponer J. Microsecond-Scale MD Simulations of HIV-1 DIS Kissing-Loop Complexes Predict Bulged-In Conformation of the Bulged Bases and Reveal Interesting Differences between Available Variants of the AMBER RNA Force Fields. *The Journal of Physical Chemistry B*. Dec.2015 119:15176–15190. <http://dx.doi.org/10.1021/acs.jpcb.5b08876>. [PubMed: 26548477]
164. Ennifar E, Dumas P. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *Journal of Molecular Biology*. 2006; 356:771–782. [PubMed: 16403527]
165. Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nature Structural & Molecular Biology*. 1998; 5:432–436.
166. Girard F, Barbault F, Gouyette C, Huynh-Dinh T, Paoletti J, Lancelot G. Dimer initiation sequence of HIV-1 genomic RNA: NMR solution structure of the extended duplex. *Journal of Biomolecular Structure and Dynamics*. 1999; 16:1145–1157. [PubMed: 10447199]
167. Baba S, Takahashi K-i, Noguchi S, Takaku H, Koyanagi Y, Yamamoto N, Kawai G. Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *Journal of Biochemistry*. 2005; 138:583–592. [PubMed: 16272570]
168. Kieken F, Paquet F, Brule F, Paoletti J, Lancelot G. A new NMR solution structure of the SL1 HIV-1Lai loop–loop dimer. *Nucleic Acids Research*. 2006; 34:343–352. [PubMed: 16410614]
169. Henriksen NM, Davis DR, Cheatham TE III. Molecular dynamics re-refinement of two different small RNA loop structures using the original NMR data suggest a common structure. *Journal of Biomolecular NMR*. 2012; 53:321–339. [PubMed: 22714631]

170. Sripakdeevong P, Cevic M, Chang AT, Erat MC, Ziegeler M, Zhao Q, Fox GE, Gao X, Kennedy SD, Kierzek R. Structure determination of noncanonical RNA motifs guided by ¹H NMR chemical shifts. *Nature Methods*. 2014; 11:413–416. [PubMed: 24584194]
171. Woodson SA. RNA folding pathways and the self-assembly of ribosomes. *Accounts of chemical research*. 2011; 44:1312–1319. [PubMed: 21714483]
172. Gleitsman KR, Herschlag DH. A kinetic and thermodynamic framework for the Azoarcus group I ribozyme reaction. *RNA*. 2014; 20:1732–1746. [PubMed: 25246656]
173. Golden, BL., Hammes-Schiffer, S., Carey, PR., Bevilacqua, PC. An integrated picture of HDV ribozyme catalysis. Springer; 2013. p. 135–167.
174. Pinamonti, G., Zhao, J., Condon, DE., Paul, F., Noè, F., Turner, DH., Bussi, G. Predicting the Kinetics of RNA Oligonucleotides Using Markov State Models. *Journal of Chemical Theory and Computation*. 2016. <http://dx.doi.org/10.1021/acs.jctc.6b00982>
175. Childs-Disney JL, Disney MD. Approaches to validate and manipulate RNA targets with small molecules in cells. *Annual Review of Pharmacology and Toxicology*. 2016; 56:123–140.
176. Disney MD, Winkelsas AM, Velagapudi SP, Southern M, Fallahi M, Childs-Disney JL. Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chemical Biology*. 2016
177. Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: Form, distribution, and function. *Science*. 2016; 352:1408–1412. [PubMed: 27313037]
178. Condon DE, Yildirim I, Kennedy SD, Mort BC, Kierzek R, Turner DH. Optimization of an AMBER Force Field for the Artificial Nucleic Acid, LNA, and Benchmarking with NMR of L (CAAU). *The Journal of Physical Chemistry B*. 2014; 118:1216–1228. [PubMed: 24377321]
179. Benner SA. Understanding Nucleic Acids Using Synthetic Chemistry. *Accounts of Chemical Research*. 2004; 37:784–797. <http://dx.doi.org/10.1021/ar040004z>. [PubMed: 15491125]
180. Ditzler MA, Otyepka M, Sponer J, Walter NG. Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Accounts of Chemical Research*. Jan.2010 43:40–47. <http://dx.doi.org/10.1021/ar900093g>. [PubMed: 19754142]
181. Šponer J, Šponer JE, Petrov AI, Leontis NB. Quantum chemical studies of nucleic acids: can we construct a bridge to the RNA structural biology and bioinformatics communities? *The Journal of Physical Chemistry B*. Dec.2010 114:15723–15741. <http://dx.doi.org/10.1021/jp104361m>. [PubMed: 21049899]
182. Šponer J, Banáš P, Jurečka P, Zgarbová M, Kührová P, Havrila M, Krepl M, Stadlbauer P, Otyepka M. Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *The Journal of Physical Chemistry Letters*. 2014; 5:1771–1782. [PubMed: 26270382]
183. Cieplak P, Caldwell J, Kollman P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry*. 2001; 22:1048–1057.
184. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B*. 2010; 114:2549–2564. <http://dx.doi.org/10.1021/jp910674d>. [PubMed: 20136072]
185. Baker CM, Anisimov VM, Alexander D, MacKerell J. Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model. *The Journal of Physical Chemistry B*. 2011; 115:580–596. <http://dx.doi.org/10.1021/jp1092338>. [PubMed: 21166469]
186. Šponer J, Mládek A, Šponer JE, Svozil D, Zgarbová M, Banáš P, Jurečka P, Otyepka M. The DNA and RNA sugar-phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Physical Chemistry Chemical Physics*. Nov. 2012 14:15257–15277. <http://dx.doi.org/10.1039/c2cp41987d>. [PubMed: 23072945]
187. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*. 2010; 38:D280–D282. http://nar.oxfordjournals.org/content/38/suppl_1/D280.abstract. [PubMed: 19880381]

188. Amunts A, Brown A, Bai X-c, Llcer JL, Hussain T, Emsley P, Long F, Murshudov G, Scheres SH, Ramakrishnan V. Structure of the yeast mitochondrial large ribosomal subunit. *Science*. 2014; 343:1485–1489. [PubMed: 24675956]
189. Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*. 2015; 349:1182–1191. [PubMed: 26292707]
190. Liberman JA, Suddala KC, Aytenfisu A, Chan D, Belashov IA, Salim M, Mathews DH, Spitale RC, Walter NG, Wedekind JE. Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112:E3485–E3494. [PubMed: 26106162]
191. Seetin MG, Mathews DH. Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints. *Journal of Computational Chemistry*. 2011; 32:2232–2244. [PubMed: 21509787]
192. Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM. Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. *Journal of the American Chemical Society*. 2009; 131:2541–2546. [PubMed: 19193004]
193. Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS. 3D RNA and Functional Interactions from Evolutionary Couplings. *Cell*. 2016; 165:963–975. [PubMed: 27087444]

Figure 1.

An example of RNA sequence (left), secondary structure (middle), [3, 4, 5] and tertiary structure (right). [6] The secondary structure records base-pairing associations for the sequence which, being mostly more favorable than other contacts, constrain tertiary structure. This is what is meant by the observation that RNA structure is hierarchical. One goal of computational methods is to accurately predict 3D structure from sequence.

Figure 2.

Geometry and partial charges of TIP3P water model, r_{O-H} is bond length, Θ is bond angle, r_{vdW} is van der Waals radius of oxygen atom, q_H is partial charge of each H atom, q_O is partial charge of O atom.

Figure 3.

This figure shows the refinement process that RNA force fields have undergone from the early '90s to the present. The black bars show these connections as a cladistic tree would, and their length is proportional to the number of years between publications. AMBER FF94, FF99, Parmbsc0, and FF10 are described in refs. Cornell 1995, Wang 2000, Perez 2007, and Zgarbova2011a, respectively. These record the development of the standard AMBER force field for RNA (note that more recent standard AMBER force fields for RNA are identical to FF10). Chi Yildirim and ParmTor are described in refs. Yildirim 2010, and Yildirim 2012, respectively. Chen and Garcia's force field is described in ref. Chen 2013, and is based on the hypothesis that FF99 overemphasizes stacking which they reduce by scaling down VDW radii and well depths by 95 and 80%, respectively, to agree with higher level QM (CCSD(T)) calculations. Thomas Cheatham's group hypothesized that phosphate oxygens might be incorrectly sized, based on work from the Case group,[114] and produced Amber +vdW BB and Amber+vdW. All use parameters described in ref. Bergonzo 2015. AmberPDB is the result of a new approach for force-field fitting, described in ref. Gil-Ley 2016, where targeted metadynamics (TMD)[105, 106] simulations enforce a free energy curve along degrees of freedom explicitly parameterized by the force field, whereupon those parameters are fit to the TMD free energy curve. In this case the enforced free energy curve was the log-odds of dihedral angles for α , β , ϵ , and from dinucleotides in the PDB. Amber + Aytenfisu is a refit of the dihedral parameters using a simultaneous linear regression to match QM energies calculated for an ensemble of dinucleotides amassed from the PDB and sampled from simulations, described in ref. Aytenfisu 2017. CHARMM has also offered force fields tuned for nucleic acids, with CHARMM22, 27, and 36 described in refs. Mackerell 1995, (Foloppe 2000 + MacKerell2000a), and Denning 2011, respectively.

Figure 4.

Umbrella sampling can sample across barriers in landscapes. The x-axis of these plots is a reaction coordinate, which can be any quantity that describes progress along the conformational change pathway, such as a distance, angle, or dihedral angle. The y-axis is the Gibbs free energy change. In panel A, a molecular landscape is shown, where there are two low free energy basins and a barrier between them. If this barrier is much larger than thermal energy, then molecular dynamics will be slow to cross it. In panel B, a restraint energy (red) is added to the landscape to lower the barrier between the two low free energy states. In this way, the barrier can be easily crossed, as shown in the final landscape (green). Although in theory this would work to enable sampling, in practice the determination of the height and shape of the barrier is the goal of the simulations, hence it is not clear what potential would suffice to use as the restraint. Panel C shows that a series of umbrella potentials (red) can be used to guide the molecule across the barrier. Each restraint is used in a separate, equilibrium simulation, and therefore each independent equilibrium simulation will sample a distinct region of the reaction coordinate. In panel D, each simulation has been analyzed to determine the free energy for each equilibrium simulation, but an arbitrary offset results for each. To recover the original landscape (panel A), the free energies are “stitched” together using WHAM.[124]

Figure 5.

Free energy integration. This example shows the alchemical transformation of a G-C pair to an iG-iC pair, where the amino and carbonyl groups are transposed relative to G and C. In free energy integration, the system is transformed from one molecule to another, where λ indicates the progress. In this example, $\lambda = 0$ is the G-C pair and $\lambda = 1$ is the iG-iC pair. For intermediate values of λ , the pair is partially G-C and partially iG-iC.

Figure 6.

(a, b) GU pairs with two and one hydrogen bond. (c, d) Sheared (trans Hoogsteen/Sugar edge A–G) and imino (cis Watson-Crick/Watson-Crick A–G) GA pairs, respectively. (e) Watson-Crick AU pair. (f) Hairpin used for umbrella sampling to predict change in G when one or both of circled GC pairs are changed to AU.

Figure 7.

Comparison between experimental and predicted change in G° (kcal/mol) at 300 K for duplex formation when GC pairs are replaced by iGiC[41]. The thermodynamic cycle requires that $\Delta G_4^0 - \Delta G_2^0 - \Delta G_1^0 + 2\Delta G_3^0 = 0$ so that $\Delta G_4^0 - \Delta G_1^0 = \Delta G_2^0 - 2\Delta G_3^0$. Revisions to FF99 for parameters of torsions, χ , α , γ , ϵ , ζ , and β increased agreement between experiment and prediction by about 10 kcal/mol.

Figure 8.

NMR structures and free energy increments at 37 C for internal loops in (5' GCUGAGGCU)₂, [148] (5' GCGGAUGCU)₂, [148] (5' GCGGACGC)₂, [149] and (5' GGCGAGCC)₂, [150]. Note that the NMR spectra do not eliminate the possibility of up to 10% of a minor structure. Values for loop ΔG° are modified from previously published values because non-loop nearest neighbors were subtracted using parameters from Xia et al. [132] and Chen et al. [133].

Figure 9.

Results from tests of different force fields used to predict structure of UUCG tetraloop.[115]

Cluster 5 agrees best with the NMR structure, [151] which is shown in green. Reproduced from ref. 115 with permission from Thomas E. Cheatham.

Figure 10.

Two structures commonly generated in MD simulations of 5'GACC. [110, 160] (Left) An A-form like structure consistent with NMR spectra. (Right) A structure with G1 intercalated between C3 and C4; NMR does not detect this structure.