Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

# Modeling nucleic acids

Adelene YL Sim[1], Peter Minary[2] and Michael Levitt[2]

Nucleic acids are an important class of biological macromolecules that carry out a variety of cellular roles. For many functions, naturally occurring DNA and RNA molecules need to fold into precise three-dimensional structures. Due to their self-assembling characteristics, nucleic acids have also been widely studied in the field of nanotechnology, and a diverse range of intricate three-dimensional nanostructures have been designed and synthesized. Different physical terms such as base-pairing and stacking interactions, tertiary contacts, electrostatic interactions and entropy all affect nucleic acid folding and structure. Here we review general computational approaches developed to model nucleic acid systems. We focus on four key areas of nucleic acid modeling: molecular representation, potential energy function, degrees of freedom and sampling algorithm. Appropriate choices in each of these key areas in nucleic acid modeling can effectively combine to aid interpretation of experimental data and facilitate prediction of nucleic acid structure.

**Addresses**
[1] Department of Applied Physics, Stanford University, Stanford, CA 94305, USA
[2] Department of Structural Biology, Stanford University, Stanford, CA 94305, USA

Corresponding author: Levitt, Michael (michael.levitt@stanford.edu)

**Current Opinion in Structural Biology** 2012, **22**:273–278

This review comes from a themed issue on
Nucleic acids
Edited by Jamie Williamson and Jody Puglisi

Available online 25th April 2012

0959-440X/$ – see front matter


http://dx.doi.org/10.1016/j.sbi.2012.03.012

## Introduction

Understanding nucleic acid structure provides essential insight into functional roles fundamental to molecular biology. Unfortunately, experimentally determining RNA and DNA structures at high resolution is tedious, expensive and not always tractable, particularly for large complex systems, which can show significant molecular flexibility.

A complementary approach is to study nucleic acid structure *in silico*. With recent discoveries of RNA's various gene regulation roles and the potential applications of nucleic acid nanostructures to biocomputing [1] and nanotechnology [2,3], there has been extensive research into developing computational tools for modeling and manipulating nucleic acid structure (Table 1; also recent reviews [4,5••] and references therein).

Since computational modeling can be used to address a wide range of problems that vary in complexity and resolution (time/size/precision), an appropriate choice of algorithm or modeling platform is required. We discuss four main aspects of nucleic acid modeling here: firstly, molecular representation; secondly, potential energy function; thirdly, degrees of freedom; and fourthly, sampling algorithm. We focus on general modeling techniques and strategies that are applicable to an array of modeling purposes such as generating an ensemble of plausible molecular models, identifying a native-like molecular structure, studying folding kinetics, probing the effects of base mutations, refining molecular models and modeling with limited experimental data.

## Molecular representation

Modeling nucleic acids with an all-atom representation and consequent precision is computationally expensive (Figure 1), particularly for large systems. Analogous to proteins, coarse-graining nucleic acids is a common approach to handle larger molecular systems [6]. Each nucleic acid base can be represented by a subset of atoms or pseudo-atoms (e.g. [7–9] for DNA and [10•,11–13,14•] for RNA; see Figure 1). Alternatively, each base can be represented as a plane [15] and modeled by rigid body parameters (Figure 1). These different coarse-grained representations have been successfully implemented in RNA structure prediction [10•,11,12].

For very large nucleic acid systems with many nucleotides, modeling can still be intractable even with such extensive coarse-graining. Because base-paired helices of DNA and RNA are rigid compared to single strands of bases, another coarse-graining strategy is to depart from atomic-level detail and instead represent each nucleic acid helix as a cylinder (Figure 1). Clearly this coarse-graining is only applicable for exploring global conformations rather than for extracting fine molecular information.

What molecular representation is most suitable? For practical considerations, it makes most sense to use the minimum possible representation that is still able to capture the phenomenon of interest with reasonable computational cost (Figure 1). While all-atom molecular modeling is usually favored due to the high level of structural detail, the cost could become astronomical

**Table 1**

**Summary of molecular modeling software**

| Modeling software | Purpose | Representation | Potential energy function | Degrees of freedom | Sampling algorithm |
|---|---|---|---|---|---|
| Amber; Nucleic Acid Builder [50] | Dynamics; Sampling | All-atom | Physics-based force-fields | All-atom | Molecular dynamics |
| BARNACLE [28] | Structure prediction | All-atom | Dihedral angle Bayesian network | Dihedral angles | Markov chain model |
| FARNA/FARFAR [25,26••] | Structure prediction | All-atom | Knowledge-based | Fragments | Monte Carlo |
| iFoldRNA [10•,17] | Structure prediction | Coarse-grained (all-atom reconstruction available) | Knowledge-based | All coarse-grained atoms | Discrete molecular dynamics |
| MacroMolecule Builder (formerly RNABuilder) [11] | Optimization | All-atom | Knowledge-based and experimental constraints | User defined | Relaxation with changing degrees of freedom and constraints (user-defined); simulated annealing |
| MC-Sym [38••] | Structure prediction | All-atom | Chain connectivity | Fragments | Cyclic building |
| MOSAICS [37] | Sampling; Optimization | Any | Any | Any/User defined; embedded degrees of freedom allowed | Monte-Carlo; Monte Carlo Minimization |
| NAMD [51] | Dynamics | All-atom | Physics-based force-fields | All-atom | Molecular dynamics |
| NAST [12] | Sampling | Coarse-grained | Knowledge-based and experimental constraints | All coarse-grained atoms | Molecular dynamics |
| Vfold [13] | Structure prediction | All-atom | Physics and knowledge-based | Fragments | Piece-wise assembly |
| YUP [52] | Sampling; Optimization | Any | Knowledge-based or user-defined | User defined | Monte Carlo; molecular dynamics; gradient-based minimization |

for large systems, making coarse-graining an absolute necessity. Perhaps an appropriate compromise would be to coarse-grain selected regions, and model important interacting sites at all-atom representation, however such a choice would require development of an interaction potential to traverse different molecular representations. Alternatively, one could also retain a high level of detail in the molecular representation, but decrease the number of degrees of freedom (DOFs) that are used in the sampling process (see 'Degrees of freedom' section). A less straightforward process is to first model with a coarse-grained representation, then restore all-atom resolution based on known or idealized nucleotide geometry [13,16,17].

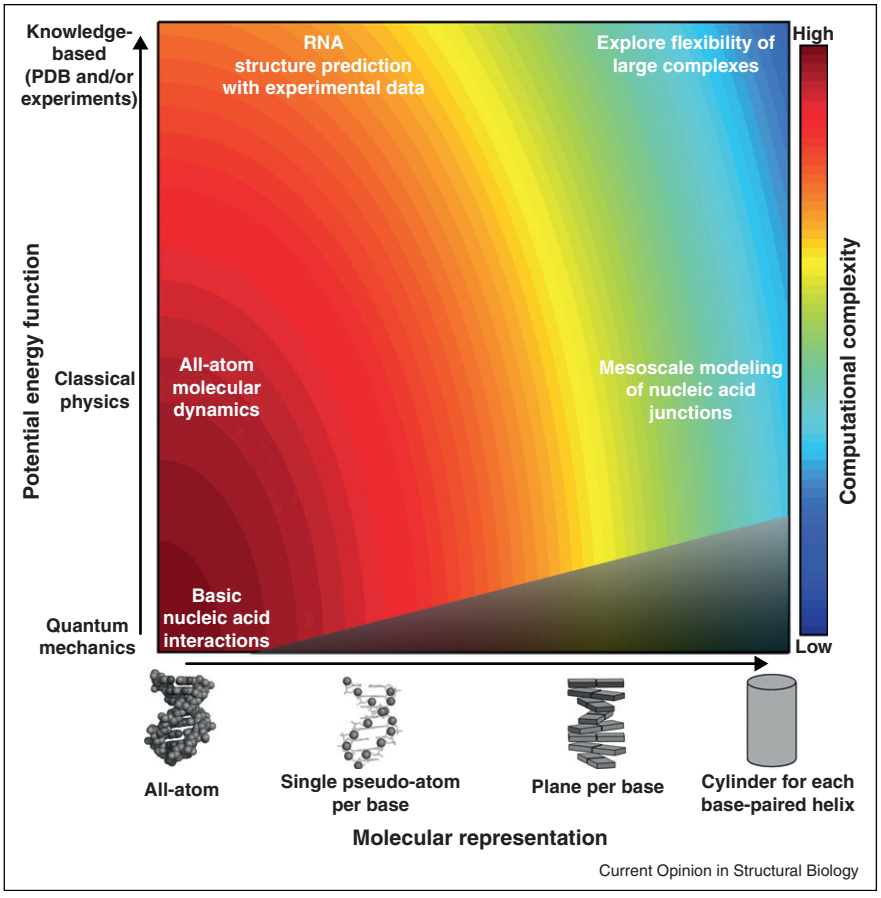
It is also important to note that here we have only discussed different representations of the nucleic acid structure, not those of its surrounding solvent or counter-ions, which can be expressed implicitly and/or explicitly. Nucleic acids carry high negative charges and hence solvent and ions must be correctly treated in order to accurately model RNA and DNA. Discussion of nucleic acid solvation and electrostatics is beyond the scope of this review and the reader is referred to other relevant literature (e.g. [18,19]).

## Potential energy function

In molecular modeling, a potential energy function is required to distinguish physical and biologically relevant conformations. These potentials come in varying degrees of precision and complexity (Figure 1), ranging from a primitive function considering only steric restraints to one that depends on quantum mechanical calculations. The choice of potential is critical to the efficacy of molecular modeling: the accuracy of modeling depends on the correctness of the potential whereas sampling efficiency varies with the nature of the potential guiding the simulation (discussed in 'Sampling algorithm' section).

Traditional physics-based potentials such as AMBER [20] and CHARMM [21] approximate atomic-level interactions as bonded (bond, bend-angle stretching, torsional rotations) and pairwise non-bonded terms, each described by analytical mathematical formulas. Parameters for these mathematical expressions were optimized using experimental observations on small molecules or else by fitting to quantum chemical calculations. A significant advantage of such physics-based potentials is their applicability to molecular dynamics simulations (see [22,23] and references therein).

**Figure 1**




This schematic diagram classifies nucleic acid modeling according to molecular representation and potential energy function; both affect computational complexity. Very detailed potentials are not applicable to molecular representations that are too coarse (dark shaded triangle). Computational complexity also depends on the number and type of degrees of freedom of the particular system that in turn depend on the potential energy function and molecular representation. Large systems with more atoms are generally more complex than small systems.

Coarse-grained systems are usually modeled with knowledge-based potentials that are typically derived from structural information gathered from high-resolution experimental structures [10•,11,12,24], from molecular models produced by all-atom simulations with physics-based potentials [9], or both [14•]. A major advantage of knowledge-based potentials is the ability to tune potentials to the particular molecular representation just by adjusting the information that is extracted from known structures. For instance, one could use different nucleotide representations to generate potentials based on known physical interactions such as nucleic acid base-pairing and stacking (compare refs. [10•,11,12,25,26••,27]), or from interatomic distances [24] or torsional angles [28], or a combination of the above [29]. Recently, a coarse-grained potential was developed for RNA using parameterized bonded and non-bonded terms like those in traditional physics-based potentials [14•]. Another advantage of knowledge-based potentials is that their mathematical forms can be easily adjusted. In some cases, these potentials are discontinuous; it was shown that RNA molecules can be efficiently modeled using a discretized knowledge-based potential [10•]. Additionally, because knowledge-based potentials implicitly capture solvent effects, they are less computationally intensive to use, in that water molecules and ions need not be explicitly included.

While physics-based potentials have been applied widely, knowledge-based potentials have mostly been used to select native-like structures; in some cases their accuracies supersede those using physics-based potentials [26••]. Despite successes in distinguishing native-like conformations, because knowledge-based parameters are usually derived from static experimental structures, it is unclear how accurate they would be for studying nucleic acid dynamics and folding pathways, particularly in different ionic solution conditions than those used to solve the experimental structures.

Another class of potentials can be derived from constraints that come in the form of secondary structure information or low-resolution experimental data. Such limited experimental data appears to provide sufficient constraints so that RNA structure determination can be more tractable, particularly for larger systems [11,12,30–33].

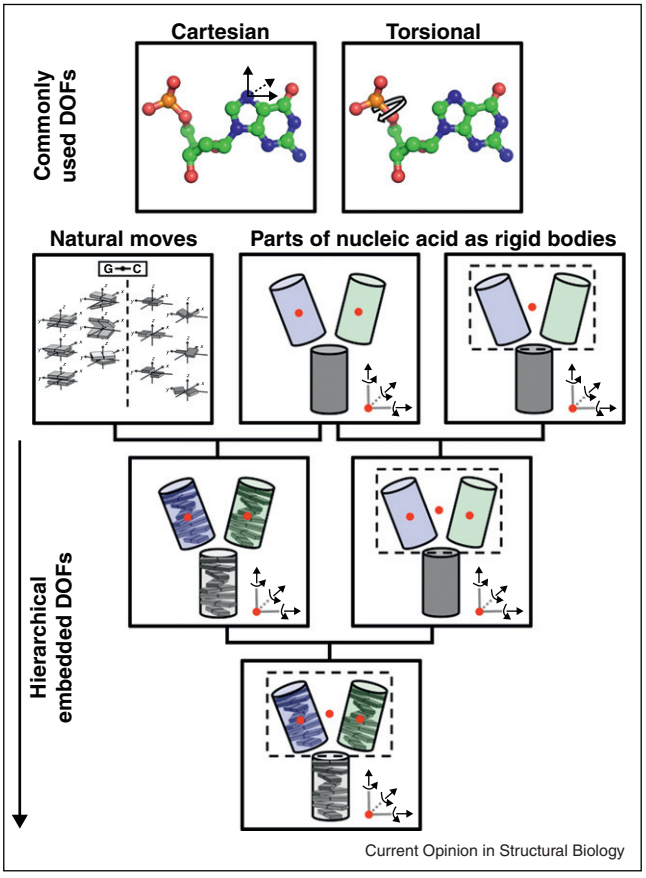
## Degrees of freedom (DOFs)

Another approach to coarse-graining large systems is to use alternate sets of DOFs while retaining an all-atom representation. By modeling using dihedral or torsional DOFs instead of all-atom Cartesian DOFs, dimensionality (number of DOFs) is substantially reduced. This technique is motivated by the limited flexibility of bond lengths and bond angles: any conformational rearrangement in Cartesian space can be accurately expressed in torsional space because bond lengths and bond angles are approximately fixed at their equilibrium values.

Unfortunately, the benefits of modeling with torsional DOFs are less obvious when modeling large and complex molecular assemblies. The root of the problem is the lever arm effect caused by rotating a dihedral angle about a single bond: a small torsional rotation about a bond can lead to substantial conformational change distant from it, rather like how a small rotation about one's elbow results in a large movement of the hand. The solution to this lever arm problem is to use torsional DOFs that only affect local conformation. However, doing this will break the continuity of the molecular chain, which needs to be repaired with a chain-closure algorithm [34]. The computational complexity of standard chain-closure algorithms makes them very expensive, encouraging the development of closure algorithms that are of reasonable computational cost. This need has been met with a linear complexity stochastic chain-closure algorithm [35••] that opens up new avenues of application: arbitrary sets of DOFs can be used, since the chain-closure algorithm can effectively restore multiple chain breakages.

Recently we applied our chain-closure algorithm to complex RNA systems, exploiting the hierarchy present in RNA structure to define different embedded sets of DOFs [36••]. The DOFs used varied in size and complexity: at the lowest level nucleotide planes move as rigid bodies with an all-atom representation; some higher level DOFs involve moving helices (base-paired regions of nucleic acids) independently or as sets of two or more helices (see Figure 2). Since these DOFs are embedded, none of the helices are perfectly rigid so that we explore a continuous phase space at the nucleotide level. In each proposed move of a Markov chain Monte Carlo step, the RNA is first regarded as a 'liquid' of individual nucleotides that are free to move based on the user-defined DOFs independent of chain connectivity. Such a complex move may lead to multiple chain breakages within the RNA that we fix with chain-closure. Using embedded DOFs (implemented in MOSAICS [37]; discussed in [36••]) allows more effective and accurate exploration of the conformational space of an RNA four-way junction than with other state-of-the-art RNA sampling methods.

**Figure 2**




Different ways of manipulating nucleic acid degrees of freedom (DOFs) in modeling. Use of Cartesian DOFs is the default in most modeling procedures, but generally results in large sampling dimensionality (number of DOFs) for systems of biomedical or nano-engineering significance. Modeling with torsional DOFs significantly reduces sampling dimensionality, however, due to the lever-arm effect (see main text), such an approach can cause a low acceptance ratio in Monte Carlo simulations. Recently, hierarchical embedded DOFs [36••] together with stochastic chain closure [35••] were introduced; both facilitate efficient sampling of nucleic acid structure. Different levels of embedded rigid body motions (centers of motions are illustrated by red dots) ensures that no segment of the nucleic acid is completely rigid, to give a sampling distribution that is similar to sampling with just natural moves but with substantially improved sampling efficiency [36••].

## Sampling algorithm

Finally, one needs to choose an appropriate dynamical/sampling/optimization algorithm to guide exploration of conformational space. The choice is primarily dependent on the question at hand. If one is interested in kinetics or folding dynamics, molecular dynamics should be used; it has been successfully employed to study folding processes

of nucleic acids (e.g. [22] and references therein). Alternatively, Monte Carlo is a general sampling algorithm that can be used to generate conformational ensembles with continuous or discontinuous potentials alike. Such ensembles can then be used to extract distributions of physical observables or to provide diverse models in structure prediction [25,38••]. If instead the objective is to locate local or global minima of the scoring energy surface (e.g. in structure refinement), optimization algorithms such as conjugate gradient minimization or simulated annealing can be used.

A major issue in molecular modeling is that sampled conformations often remain close to the initial starting model. This is because almost all potential energy functions give rise to energy barriers between low energy conformational basins; these barriers hinder transitions between different low energy states. Sometimes such barriers are unnaturally high due to the potentials and DOFs used. Hence conformational transitions occur with low probability in simulations, impeding effective sampling of conformational space [39].

In order to overcome these limitations, several technological innovations have been proposed to speed up conformational exploration using the aforementioned sampling algorithms. Perhaps the most popular approach is to implement multi-canonical sampling algorithms that manipulate either the temperature [40,41] or the potential [42,43]. Other solutions to overcome barriers in extra dimensions include data augmentation [44] and coupling to extended systems [45,46]. More complex solutions reformulate the basic partition function (that describes the system in thermodynamic equilibrium) in terms of new variables [47] or change the underlying potential surface [48,49].

## Conclusions

We have reviewed some key aspects and recent developments in the field of nucleic acid modeling. Despite improved computational resources and increased importance of modeling in nucleic acid nanotechnology and biomedicine, it remains crucial to interpret modeling results with utmost care: the accuracy of any particular molecular modeling technique is only as good as the potential energy function used. To avoid overinterpretation of modeling data, one needs to be aware of the limitations of the selected modeling protocol. Nonetheless, in many instances, carefully selected combinations of the appropriate molecular representation, potential energy function, DOFs and sampling algorithm yield sufficiently accurate modeling results that can facilitate interpretation of experimental data and/or to guide future experiments.

## Acknowledgements

We apologize to researchers for being unable to cite all their interesting and relevant work in the field of nucleic acid modeling due to space constraints. We are supported by NIH awards (GM041455 and GMN063817). ML is the Robert W and Vivian K Cahill Professor of Cancer Research. AYLS is an Agency for Science, Technology and Research (Singapore) scholar.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Qian L, Winfree E: **Scaling up digital circuit computation with DNA strand displacement cascades**. *Science* 2011, **332**:1196-1201.

2. Pinheiro AV, Han D, Shih WM, Yan H: **Challenges and opportunities for structural DNA nanotechnology**. *Nat Nanotechnol* 2011, **6**:763-772.

3. Guo P: **The emerging field of RNA nanotechnology**. *Nat Nanotechnol* 2010, **5**:833-842.

4. Laing C, Schlick T: **Computational approaches to 3D modeling of RNA**. *J Phys: Condens Matter* 2010, **22**:283101.

5. •• Laing C, Schlick T: **Computational approaches to RNA structure prediction, analysis, and design**. *Curr Opin Struct Biol* 2011:1-13.

A recent review on RNA structure prediction (secondary and tertiary). Some current three-dimensional structure prediction tools were compared and analyzed.

6. Pincus DL, Cho SS, Hyeon C, Thirumalai D: **Minimal models for proteins and RNA from folding to function**. *Prog Mol Biol Transl Sci* 2008, **84**:203-250.

7. Dans PD, Zeida A, Machado MR, Pantano S: **A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics**. *J Chem Theory Comput* 2010, **6**:1711-1725.

8. Savelyev A, Papoian GA: **Chemically accurate coarse graining of double-stranded DNA**. *Proc Natl Acad Sci U S A* 2010, **107**:20340-20345.

9. Knotts TA, Rathore N, Schwartz DC, de Pablo JJ: **A coarse grain model for DNA**. *J Chem Phys* 2007:126.

10. • Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV: **Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms**. *RNA* 2008, **14**:1164-1173.

Discrete molecular dynamics together with a coarse-grained representation of RNA is applied to predict RNA structure, and also used to model RNA folding processes.

11. Flores SC, Altman RB: **Turning limited experimental information into 3D models of RNA**. *RNA* 2010, **16**:1769-1778.

12. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB: **Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters**. *RNA* 2009, **15**:189-199.

13. Cao S, Chen SJ: **Physics-based de novo prediction of RNA 3D structures**. *J Phys Chem B* 2011, **115**:4216-4226.

14. • Pasquali S, Derreumaux P: **HiRE-RNA: a high resolution coarse-grained energy model for RNA**. *J Phys Chem B* 2010, **114**:11957-11966.

An RNA knowledge-based potential that describes bonded and non-bonded interactions using functional forms similar to those in all-atom physics-based potentials.

15. Lu XJ, Olson WK: **3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures**. *Nat Protoc* 2008, **3**:1213-1227.

16. Jonikas MA, Radmer RJ, Altman RB: **Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models**. *Bioinformatics* 2009, **25**:3259-3266.

17. Sharma S, Ding F, Dokholyan NV: **iFoldRNA: three-dimensional RNA structure prediction and folding**. *Bioinformatics* 2008, **24**:1951-1952.

18. Auffinger P, Hashem Y: **Nucleic acid solvation: from outside to insight**. *Curr Opin Struct Biol* 2007, **17**:325-333.

19. Draper DE: **RNA folding: thermodynamic and molecular descriptions of the roles of ions**. *Biophys J* 2008, **95**:5489-5495.

20. Pérez A, Marchan I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, Orozco M: **Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers**. *Biophys J* 2007, **92**:3817-3829.

21. 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**. *J Comput Chem* 2011, **32**:1929-1943.

22. Ditzler MA, Otyepka M, Sponer J, Walter NG: **Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in**. *Acc Chem Res* 2010, **43**:40-47.

23. Pérez A, Luque FJ, Orozco M: **Frontiers in molecular dynamics simulations of DNA**. *Acc Chem Res* 2012, **45**:196-205.

24. Bernauer J, Huang X, Sim AYL, Levitt M: **Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation**. *RNA* 2011, **17**:1066-1075.

25. Das R, Baker D: **Automated de novo prediction of native-like RNA tertiary structures**. *Proc Natl Acad Sci U S A* 2007, **104**:14664-14669.

26. •• Das R, Karanicolas J, Baker D: **Atomic accuracy in predicting and designing noncanonical RNA structure**. *Nat Methods* 2010, **7**:291-294.
The authors show that physics-based potentials (with implicit solvent) are not as effective in discriminating native-like models when compared to a high-resolution knowledge-based potential. A double mutant sequence of a signal recognition particle domain was predicted to stabilize the RNA relative to its native, and this result was verified experimentally.

27. Morozov AV, Fortney K, Gaykalova DA, Studitsky VM, Widom J, Siggia ED: **Using DNA mechanics to predict in vitro nucleosome positions and formation energies**. *Nucleic Acids Res* 2009, **37**:4707-4722.

28. Frellsen J, Moltke I, Thiim M, Mardia KV, Ferkinghoff-Borg J, Hamelryck T, Gardner P: **A probabilistic model of RNA conformational space**. *PLoS Comp Biol* 2009, **5**:e1000406.

29. Taxilaga-Zetina O, Pliego-Pastrana P, Carbajal-Tinoco MD: **Three-dimensional structures of RNA obtained by means of knowledge-based interaction potentials**. *Phys Rev E* 2010, **81**:041914.

30. Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB, Herschlag D: **Structural inference of native and partially folded RNA by high-throughput contact mapping**. *Proc Natl Acad Sci U S A* 2008, **105**:4144-4149.

31. 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**. *J Am Chem Soc* 2009, **131**:2541-2546.

32. Yang SC, Parisien M, Major F, Roux B: **RNA structure determination using SAXS data**. *J Phys Chem B* 2010, **114**:10039-10048.

33. Seetin MG, Mathews DH: **Automated RNA tertiary structure prediction from secondary structure and low-resolution restraints**. *J Comput Chem* 2011, **32**:2232-2244.

34. Ulmschneider JP, Jorgensen WL: **Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias**. *J Chem Phys* 2003, **118**:4261-4271.

35. •• Minary P, Levitt M: **Conformational optimization with natural degrees of freedom: a novel stochastic chain closure algorithm**. *J Comput Biol* 2010, **17**:993-1010.
Stochastic chain closure can never fail unlike its analytical counterpart. This, together with rapid correction of spoiled local stereochemistry facilitates efficient use of natural moves in macromolecular modeling.

36. •• Sim AYL, Levitt M, Minary P: **Modeling and design by hierarchical natural moves**. *Proc Natl Acad Sci USA* 2012, **109**:2890-2895.
A hierarchy of embedded move sets greatly improves Monte Carlo sampling efficiency, allowing previously intractable problems to be solved.

37. Minary P: Methodologies for Optimization and SAmpling In Computational Studies (MOSAICS), http://csb.stanford.edu/~minary/MOSAICS.html.

38. •• Parisien M, Major F: **The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data**. *Nature* 2008, **452**:51-55.
First easily accessible online server for modeling RNA three-dimensional structure from sequence alone. Secondary structure prediction (MC-Fold) precedes tertiary structure assembly (MC-Sym). Both steps utilize information from nucleotide 'cyclic motifs' (a specific class of RNA fragments) derived from RNA structures in the PDB.

39. Brooks CL, Onuchic JN, Wales DJ: **Taking a walk on a landscape**. *Science* 2001, **293**:612-613.

40. Swendsen RH, Wang JS: **Nonuniversal critical dynamics in Monte-Carlo simulations**. *Phys Rev Lett* 1987, **58**:86-88.

41. Geyer CJ: *Computing science and statistics: The 23rd Symposium on the Interface: Interface Foundation, vol. 156*; *Fairfax: 1991:163*.

42. Kou SC, Zhou Q, Wong WH: **Equi-energy sampler with applications in statistical inference and statistical mechanics**. *Ann Statist* 2006, **34**:1581-1619.

43. Curuksu J, Zacharias M: **Enhanced conformational sampling of nucleic acids by a new Hamiltonian replica exchange molecular dynamics approach**. *J Chem Phys* 2009, **130**:104110.

44. Tanner MA, Wing HW: **The calculation of posterior distributions by data augmentation**. *J Am Statist Assoc* 1987, **82**:528-540.

45. Minary P, Martyna GJ, Tuckerman ME: **Algorithms and novel applications based on the isokinetic ensemble. I. Biophysical and path integral molecular dynamics**. *J Chem Phys* 2003, **118**:2510-2526.

46. Minary P, Tuckerman ME, Martyna GJ: **Long time molecular dynamics for enhanced conformational sampling in biomolecular systems**. *Phys Rev Lett* 2004, **93**:150201.

47. Minary P, Tuckerman ME, Martyna GJ: **Dynamical spatial warping: a novel method for the conformational sampling of biophysical structure**. *SIAM J Sci Comput* 2008, **30**:2055-2083.

48. Li ZQ, Scheraga HA: **Monte-Carlo-minimization approach to the multiple-minima problem in protein folding**. *Proc Natl Acad Sci U S A* 1987, **84**:6611-6615.

49. Rahman JA, Tully JC: **Puddle-skimming: an efficient sampling of multidimensional configuration space**. *J Chem Phys* 2002, **116**:8750-8760.

50. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossvai I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA: *AMBER 12*. San Francisco: University of California; 2012.

51. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *J Comput Chem* 2005, **26**:1781-1802.

52. Tan RKZ, Petrov AS, Harvey SC: **YUP: a molecular simulation program for coarse-grained and multiscaled models**. *J Chem Theory Comput* 2006, **2**:529-540.