CONVERGENCE AND REPRODUCIBILITY IN SIMULATIONS OF NUCLEIC ACIDS

Allocation: NSF PRAC/7.0 Mnh
PI: Thomas E. Cheatham, III[1,2]
Collaborators: Christina Bergonzo[2], Rodrigo Galindo-Murillo[2], Daniel R. Roe[2]

[1]Center for High-Performance Computing, University of Utah
[2]University of Utah

EXECUTIVE SUMMARY:

Over the past 20-plus years, we have developed and used biomolecular simulation methods to provide insight into nucleic acid structure and dynamics. Along the way, we have faced significant challenges in the form of force-field and sampling limitations. With access to resources like Blue Waters, we have reached the tipping point where we can fully sample ensembles of small RNA molecules and DNA helices and, with recent force-field modifications, are able to properly sample RNA conformational distributions in agreement with experiment. This opens the door for the simulations we really want to perform; rather than simple assessment and validation of simulation results, we can use these powerful biomolecular simulation methods to provide significant insight into functional RNA structure and dynamics in riboswitches, ribozymes, interfering RNA, and protein–nucleic acid interactions.

INTRODUCTION

As more is understood about the biological importance of RNA, the ability to peer into this biomolecule's dynamics on an all-atom level is of increasing interest. For example, it is known that RNA molecules (riboswitches) can detect the concentration of certain metabolites leading to structural changes that alter import or production of the metabolite. RNA is also involved in gene regulation, protein production, and many emerging functional roles. Keys to RNA's ability to function are not only its structure and dynamics, but also conformational rearrangements induced by changes in its environment. Understanding the detailed atomic structure and dynamics of RNA provides insight into how RNA functions and potentially how to modulate this function, for example through drug binding.

As atomic-resolution experimental approaches have difficulty in resolving dynamic structures or structures that populate multiple configurations, there is a need to employ alternative approaches to characterize RNA structure and dynamics. High-level molecular dynamics (MD) simulations provide an additional means to study RNA dynamics (and therefore its function) with atomic-level detail. However, for such simulations to be useful they need to not only capture all of the bio-relevant conformations, but also sample correct structure populations. The former depends on the accuracy of the parameters used during the simulation (referred to as a "force field"), while the latter is a challenging problem requiring a large amount of aggregate simulation time. These two issues are related: unless you have relatively complete sampling, it is difficult to determine if inaccuracies are the result of poor parameters or if they result from limited sampling. Until recently, the challenge in simulating RNA has been validating putative changes intended to improve RNA force fields that have been generated with limited sampling of RNA structure and dynamics. Using Blue Waters, we have been able to divorce the limitation of sampling from the dynamics and structures described by the force fields, allowing a totally unbiased look at each force field's description of both DNA and RNA [1–5] .

METHODS & RESULTS

By using multi-dimensional replica exchange MD (M-REMD), we have shown that we can reproducibly and efficiently generate effectively complete structure ensembles for a variety of RNA motifs. In previous simulations we found that a particular tetranucleotide, rGACC, sampled a wide range of conformations, though experimental nuclear magnetic resonance spectroscopy results indicated only two major structures were present [6,7]. Using Blue Waters, we have expanded testing of current force fields to include combinations of non-bonded parameter modifications with several ion and water models.

The results indicate that with improved parameters, including a modified van der Waals radii set [8] and improved water model (OPC) [9], the ensemble of rGACC structures generated by M-REMD are now in quantitative agreement with experimental nuclear magnetic resonance. Comparison between two independent runs shows the high level of convergence achievable only through use of Blue Waters.

More recent work on the UUCG tetraloop structure (which is a small helical RNA stem with a r(UUCG) sequence loop on top), where it has been problematic for all current force fields to capture the native structure, shows that including these modifications in the presence of the improved OPC water model increases the percent of native structure found by an order of magnitude. Due to the high amount of sampling required to converge these simulations, this study is ongoing.

WHY BLUE WATERS?

A key innovation enabling our work is the implementation of AMBER on NVIDIA GPUs. This results in the fastest molecular dynamics code available to date [10]. Improvements in the past year have included the ability to repartition the solute's hydrogen masses, allowing us to double the simulation time step and achieve an almost instant two-fold increase in simulated time vs. real time [11]. Combined with the M-REMD framework, this very effectively reduced, by orders of magnitude, the real-world time needed to obtain the converged ensemble. With larger ensembles and combined ensembles enabled by the next generation of Track-1 systems, we will be able to move into the detailed investigation of more bio-relevant nucleic acid structures including riboswitches, ribozymes, and various protein–nucleic interactions with full reproducibility and convergence of the results.

PUBLICATIONS

Galindo-Murillo, R., D. R. Roe, and T. E. Cheatham, 3rd, Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta,* 1850:5 (2015), pp. 1041–1058, doi:10.1016/j.bbagen.2014.09.007.

Bergonzo, C., et al., Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.,* 2014. 10:1 (2014), pp. 492– 499, doi:10.1021/ct400862k.

Galindo-Murillo, R., D. R. Roe, and T. E. Cheatham, 3rd, On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nature Commun.,* 5 (2014), 5152, doi:10.1038/ncomms6152.

Roe, D. R., C. Bergonzo, and T. E. Cheatham, 3rd, Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J. Phys. Chem. B,* 118:12 (2014), pp. 3543– 52, doi:10.1021/jp4125099.