HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:

Curr Opin Struct Biol. 2024 August ; 87: 102847. doi:10.1016/j.sbi.2024.102847.

Embracing exascale computing in nucleic acid simulations

Jun Li[a], **Yuanzhe Zhou**[a], **Shi-Jie Chen**[a,*]

[a]Department of Physics, Department of Biochemistry and Institute for Data Science and Informatics, University of Missouri, 223 Physics Bldg., Columbia, 65211, MO, USA

Abstract

This mini-review reports the recent advances in biomolecular simulations, particularly for nucleic acids, and provides the potential effects of the emerging exascale computing on nucleic acid simulations, emphasizing the need for advanced computational strategies to fully exploit this technological frontier. Specifically, we introduce recent breakthroughs in computer architectures for large-scale biomolecular simulations and review the simulation protocols for nucleic acids regarding force fields, enhanced sampling methods, coarse-grained models, and interactions with ligands. We also explore the integration of machine learning methods into simulations, which promises to significantly enhance the predictive modeling of biomolecules and the analysis of complex data generated by the exascale simulations. Finally, we discuss the challenges and perspectives for biomolecular simulations as we enter the dawning exascale computing era.

1. Introduction

The emergence of exascale computing capable of quintillion (10^{18}) calculations per second marks a new era in high-performance computing (HPC) for biomolecular simulations, driven by the unparalleled computational prowess of supercomputers such as Frontier from the United States, OceanLight and Tianhe-3 from China, and the anticipated arrival of Aurora and El Capitan in the United States, as well as JUPITER in Europe. The pre-exascale computing has enabled biomolecular simulations that encompass time scales from femtoseconds to milliseconds and system sizes that can accommodate large macromolecular complexes [1]. The atomistic level of detail is essential for enhancing our understanding of biological processes, from protein and RNA folding to the biomolecular interactions within cellular environments. For example, Galvanetto et al. simulated biomolecular condensate dynamics for millions of atoms over 6.02 microseconds, taking approximately 6 months of supercomputer time [2]. By simulating systems from drug-target interactions to large viral particle assembly, HPC community made significant contributions to fight the COVID-19 pandemic [3]. The transition from pre-exascale computing to exascale computing may facilitate new breakthroughs in biomolecular simulations [4, 5]. For example, the enhanced sampling methods for exascale simulations have advanced the classification of phosphatase

* chenshi@missouri.edu (Shi-Jie Chen).

Declarations of interest: None

Declaration of generative AI in scientific writing: None

and tensin homolog (PTEN) missense variants of uncertain significance in PTEN-related cancers [6].

In the emerging era of exascale computing, the utilization of this advanced technology for biomolecular simulations, particularly in studying nucleic acids, is beginning to be explored. Nucleic acids are crucial for cellular functions at the level of transcription, translation and the regulation of gene expression. The implementation of exascale computing is expected to enable simulations of nucleic acid systems with unprecedented temporal and spatial resolution. This mini-review aims to explore the recent advances of bimolecular simulations in the nascent exascale era, especially for nucleic acids. Fig. 1 presents the pertinent hardware and software components that are involved in exascale computing for biomolecular simulations with potential applications, and subsequent sections will elaborate on specific aspects.

2. Biomolecular simulations with next-generation supercomputers

Biomolecular simulations have been profoundly revolutionized by the advent of novel computer architectures, such as Anton 3 [7, 8], MODYLAS [9], Alkaid [10], and Folding@home [11]. The novel computing architectures have offered unprecedented computational speeds and capabilities in handling complex biomolecular systems.

Anton 3 [7, 8] is a specialized supercomputer designed for molecular dynamics (MD) simulations with remarkable speed and efficiency, which adopts an advanced chip design that leverages increased parallelism and high-speed communication channels. It can perform multi-microsecond simulations for systems of more than 50 million atoms that are crucial for scientific discovery and drug development.

MODYLAS [9] is an exascale high-performance MD simulation program deployed on the Fugaku supercomputer. MODYLAS features highly efficient communication, and maintains performance even with high parallelization. It can efficiently handle large systems, such as a 101.8-million-atom system, distributing data across 32,768 nodes in 2.3 ms per MD step.

Alkaid [10] is an advanced HPC cluster specifically engineered to deliver breakthrough performance in MD simulations at the microsecond scale per day for systems comprising millions of atoms. Notably, Alkaid achieves this high-level performance with significantly reduced power consumption, making it an environmentally conscious alternative to traditional supercomputers.

Folding@home [11] represents a novel approach in biomolecular simulations, utilizing a massively parallel computing strategy through global volunteer resources. Remarkably, it has evolved into the world’s first exascale computing resource for studying the SARS-CoV-2 virus [11]. Through the Folding@home distributed computing platform, over a million citizen scientists contributed to simulating the viral proteome for 0.1 seconds, providing insights into the dramatic opening of the apo spike complex and revealing over 50 cryptic pockets that may be targeted for antiviral drug design [12].

The advancement of the above supercomputer architectures highlights a trajectory towards specialized, energy-efficient computation with massive parallelism and high-speed communication, enabling breakthroughs in scientific discovery and drug development [13]. The continued evolution of these architectures promises to address current scaling challenges and further enhance our computational capabilities in the exascale era.

3. Force fields for nucleic acids

MD simulations provide detailed insights into biomolecules, achieving high spatial and temporal resolution beyond the reach of conventional experimental approaches. The efficacy of these simulations is contingent on the precision of the force fields employed. However, current force fields rely on Lennard-Jones (LJ) and electrostatic parameters that have not been optimized for nucleic acids [14]. Moreover, the structures and dynamics of nucleic acids are influenced by the choice of force fields, such as for the DNA mini-dumbbell system [15].

While no force field is without imperfections, continuous advancements have been achieved in refining the force fields for nucleic acids. The Shaw group refined the AMBER ff14 RNA force field to more accurately represent nucleobase stacking, base pairing, and torsional conformers in RNA structures [16]. Subsequently, they optimized the DNA force field parameters to better characterize the thermal stability and conformational flexibility of single- and double-stranded DNA structures [17]. Tumuc1, a new DNA force field derived from quantum calculations [18], offers improved simulation accuracy for various DNA structures and folding processes, enhancing the understanding of DNA structures and dynamics through MD simulations. Mlýnský et al. proposed a modification to the AMBER OL3 force field by adjusting the nonbonded parameters for the intranucleotide base-phosphate interaction to fix steric clashes in RNA nucleotide structures, yielding better alignment with experimental data for various RNA motif structures [19].

In contrast to the traditional force fields, a polarizable force field offers the capability of dynamically altering the electron distributions within biomolecules, thus more precisely replicating the interactions and dynamics. However, limitations still exist for polarizable force fields. For example, the current Drude polarizable force field can not accurately model the single stranded RNA structures [20] and the duplex DNA structures [21]. Moreover, simulations in polarizable force fields are generally slower than in traditional (non-polarizable) force fields due to the additional complexity and calculations. Consequently, the advent of exascale computing presents a promising avenue for the efficient application of polarizable force fields, necessitating further optimization of the calculations. Furthermore, in the exascale computing era, the long-time simulations of large-scale systems can in turn facilitate the validation of the accuracy of the force fields for nucleic acids.

In addition to the accuracy of force fields for nucleic acids, the transition from petascale to exascale computing requires new algorithms for force calculation that can exploit the vast parallelism for the new supercomputers [4, 5]. For most systems, the computation of electrostatic interactions proves to be the most time-intensive and its accuracy is critical for modeling the behavior of biomolecular systems [22]. The recent ANKH method [23]

emerges as a novel solution, applying an interpolated Ewald strategy to achieve O(N) computational complexity. Comparatively, the ANKH method maintains the accuracy seen in the traditional methods while offering better scalability and reduced computational demands. This makes it particularly advantageous for modern HPC architectures, where it can leverage parallelization and GPU capabilities to accommodate the increasing size and complexity of biomolecular simulations.

4. Enhanced sampling methods

In the pursuit of harnessing exascale computing capabilities of massively parallel CPUs and GPUs, it is also essential to develop enhanced sampling techniques tailored to this advanced scale. Various enhanced sampling methods have significantly improved our ability to study rare or slow events in molecular systems via MD simulations [24], such as umbrella sampling (US), metadynamics (MetaD), replica exchange molecular dynamics (REMD), steered molecular dynamics (SMD), adaptive biasing force (ABF), temperature accelerated molecular dynamics (TAMD). For example, Sanbonmatsu and colleagues applied MetaD to investigate the energy landscape of the SAM-I riboswitch using a modified collective variable (CV) based on tertiary contacts, and validated their simulations through SAXS experiments [25]. A recent work involving the integration of runtime steering in MD simulations presents a paradigm shift from post-simulation analysis to active, on-the-fly analysis [26]. This *in situ* framework enables the early termination or restart of simulations based on the real-time analysis of CVs, leading to a more rapid and resource-efficient exploration of conformational space. Unlike the centralized data analysis methods, which scale poorly on large datasets, the *in situ* framework's distributed nature enables scalable and efficient MD simulations on HPC systems.

5. Coarse-grained modeling of nucleic acids

Compared with all-atom models, coarse-grained (CG) models for nucleic acids have far fewer atoms or beads and a much smoother energy landscape, enabling more efficient conformational sampling. Various CG models have been employed in nucleic acid research. Some focus on structure prediction, such as SimRNA [27], RNApps [28], IsRNA [29], RNAJP [30] for RNA 3D structure prediction and Shi's model [31] for DNA 3D structure prediction. Some other CG models focus on studying the thermodynamic and physical properties, and folding processes of nucleic acid structures, such as HiRe-RNA [32] and oxRNA [33] for RNA molecules, oxDNA [34] for DNA molecules. For a more extensive list, please refer to the provided references [35, 36]. CG models excel in efficiency when applied to large systems, such as chromatin, chromosomes, and nanostructures composed of DNA or RNA. We will present selected recent studies employing CG models to investigate these sizable structures.

Li et al. performed Brownian dynamics simulations at the nucleosome resolution with GPUs to study mesoscale chromatin fibers [37]. The results from these simulations align well with experimental observations, providing insights into the behavior of chromatin under various conditions, including the presence of different salt concentrations and linker histones. This

study is instrumental in enhancing the understanding of chromatin's role in gene regulation and related diseases.

Wasim et al. developed a data-driven CG model to study the *E. coli* chromosome of 4.6×10^6 base pairs (bp), integrating the harmonic distance restraints converted from the genomic contact probability map at 5000-bp resolution derived via the Hi-C experiment, and the annotation of chromosome regions at 500-bp resolution obtained from RNA-sequencing data [38]. In their model, 500-bp DNA is reduced as a bead and the unreplicated chromosome is characterized as a polymer of 9,280 such beads. Through MD simulations, this polymer-physics based model produced a conformational ensemble that the *E. coli* chromosome can adopt under the experimental constraints, demonstrating that it spontaneously forms distinct macrodomains. The authors also extended the model to study the replicating chromosome. Prior to this research, Lappala et al. had developed a method known as "4DHiC" [39], which incorporated the harmonic distance restraints derived from Hi-C contact maps evolving over time into CG simulations. They used this method to spatiotemporally study the structural reorganization of the mammalian X chromosome at 200-kb resolution.

Tan et al. developed residue-level CG models for protein, RNA, and DNA for large-scale MD simulations, which can be efficiently implemented in the software GENESIS [40]. They optimized the algorithms and enhanced the parallelization performance on HPC. Using the new models, they studied the large-scale systems including the virus capsid of 1,687,980 CG beads and the chromatin of 2,318,796 CG beads corresponding to 1024 nucleosomes and 219,213 DNA bp.

Wong et al. used the nucleotide-level CG model oxDNA in conjunction with umbrella sampling to delineate the free-energy landscapes pertinent to the mechanical deformation of large-scale DNA nanostructures [41]. This methodology holds significant promise for advancing the field of DNA mechanotechnology and in elucidating the stress dynamics during the self-assembly of DNA origami structures. Moreover, an advanced and convenient graphical web service for conducting, visualizing, and analyzing oxDNA and oxRNA MD simulations can be accessed at OxDNA.org, accelerated by GPU-optimized HPC servers [42].

While previous studies have affirmed the efficacy of computational simulations employing CG models for investigating large-scale nucleic acid systems, it is important to note that these simulations may encounter limitations in terms of accuracy. These limitations arise from the inherent coarse-graining of nucleic acids and the omission of crucial environmental factors, such as solvent and ions, which can impact the fidelity of the simulations. In the forth-coming era of exascale computing, we anticipate the ability to incorporate finer details into large-scale simulation systems, allowing for longer and more comprehensive simulations.

6. Simulations of nucleic acids with ligands

RNA-targeted drug discovery is an emerging field for designing novel drugs. Though structure-based virtual screening (SBVS) using molecular docking has proven useful for

identifying potential lead compounds for various RNA targets [43, 44], MD simulations offer more accurate predictions of structural, thermodynamic, and kinetic properties albeit at greater computational cost. Here we highlight selected recent MD studies focused on characterizing RNA-ligand interactions, demonstrating the growing potential and need for enhanced computing resource of this versatile technique as we enter exascale computing era.

Paternoga et al. performed all-atom explicit-solvent MD simulations to investigate the temperature influence (−180 to 37°C) on bound drug (lincomycin) and water molecules for a large ribosomal subunit (LSU) [45]. Their results revealed that while all water molecules remained stably bound at 37°C, increased fluctuations occurred at higher temperatures. They also found the extent of the fluctuations varied for different water molecules in a manner that aligned well with the cryo-EM density map. This suggests a viable strategy for designing new antibiotic derivatives by identifying regions within the lead compounds that could be modified to displace stably bound waters and interact with the RNA target.

Wang et al. studied the dissociation of cognate and synthetic ligands from a *Tte*-PreQ$_1$ riboswitch aptamer system at an all-atom resolution through an artificial intelligence (AI)-augmented biased MD simulation method [46]. To tackle the challenge of finding a proper *a priori* estimate of the reaction coordinates (RCs), Wang et al. employed an AI-based sampling method to automate the learning of the dominant slow degrees of freedom on the fly. This scheme iterates between typical MD to generate data and deep learning to construct approximate RCs from the data. The optimized RCs were expressed as a linear combination of different RNA-ligand heavy atom contacts and were used to run several independent biased MD simulations using well-tempered metadynamics to estimate the free energy profiles of the systems. The AI-augmented simulations indicated that cognate and synthetic ligands have different preferences for two observed dissociation pathways. Furthermore, the simulations predicted that mutations of two nucleotides distal to the binding site, which exhibited the most significant relative movement during the dissociation process, would have differing impacts on the cognate and synthetic ligand-bound systems. Subsequent mutagenesis experiments measuring equilibrium dissociation constant K_D for six mutant RNA-ligand systems validated these predictions. This work demonstrates a novel AI-assisted strategy to access the slower timescales of ligand dissociation, which are typically inaccessible to standard MD simulations.

The implementation of hardware and software infrastructures to enable exascale computing is anticipated to rapidly advance the field [13, 47]. Such platforms not only allow the extension of accessible system sizes, timescales, and accuracy of classical MD methods, but also facilitate a wider use of hybrid approaches such as quantum mechanics/molecular mechanics (QM/MM) simulations.

7. Integration of machine learning in biomolecular simulations

Machine learning tools, such as AlphaFold [48], have been heavily employed for protein structure prediction, benefiting from the ability to associate complex patterns in large datasets to predict protein folds from sequences. The success of machine learning applications in structural biology is underscored by their notable accuracy in predicting

protein structures within CASP competitions [49], although, for RNA structure prediction, machine learning approaches lag behind conventional methods [50, 51]. Despite the high degree of precision in protein structure prediction accuracy afforded by current machine learning methodologies, they are limited to yielding a singular, static conformation. This constraint renders them inadequate for elucidating the array of transitional structures pertinent to the execution of biological functions. Moreover, these methods do not furnish insights into dynamic properties, thereby rendering MD simulations an essential complement for comprehensive structural analysis [52].

The integration of machine learning with exascale computing is revolutionizing biomolecular simulations for protein and nucleic acid folding and dynamics. The development of force fields is undergoing a transformation due to machine learning, which can predict a wide range of functional forms for energies and forces, thus capturing complex multi-body correlations that may be missing in classical force fields [53, 54]. QM force fields are instrumental in the detailed modeling of electronic structures pertinent to drug design and enzymatic catalysis. However, biomolecular computations utilizing QM are notably time-intensive. Recent advancements have seen machine learning methods substantially enhance the accuracy and efficiency of calculating QM force fields [55]. Similarly, modeling protein in coarse-grained force fields is also being developed using machine learning to treat larger systems and longer timescales [56]. For RNA, there exist some machine learning-based scoring functions [57, 58]. However, they are generally used to score RNA 3D structures and can not be used for MD simulations. Therefore, the development of machine learning-based force fields for RNA is an area that is yet to be fully explored and realized. Recently, the latest version 8 of the MD package OpenMM has been released, which can perform MD simulations with machine learning potentials [59].

Furthermore, advancements in computing power have significantly enhanced the capabilities of long MD simulations. This increase in data generation necessitates the development of machine learning techniques to make sense of these vast datasets. Machine learning is making a significant impact in the analysis of protein dynamics [60, 61]. The future is anticipated to see an increased application of machine learning techniques in the analysis of extensive data for nucleic acid systems and the development of interpretable machine learning models [62].

These advancements resulting from machine learning methods are likely to continue evolving, providing deeper insights into the mechanisms underlying protein/nucleic acid-related diseases and the discovery of new therapeutic approaches. While machine learning offers powerful tools for advancing biomolecular sciences, it does not replace but rather complements the invaluable physical and chemical knowledge necessary to design and interpret complex biomolecular simulations.

8. Challenges and perspectives

While the power of exascale systems is formidable, it brings with it the need for advancements in software development, data management, and methodological approaches that can fully utilize this power. One of the main challenges lies in the integration of

these powerful computing systems into the existing framework of biomolecular simulation. Despite the advances, there is still a need for extensive programming and coding efforts to adapt or develop new simulation codes that can efficiently harness the computational might of exascale computers. The complexity of this task is compounded by the heterogeneous architecture of exascale systems, requiring a sophisticated understanding of how to best allocate computational tasks to diverse processing units. Another challenge is to ensure that the simulation results are not only accurate but also relevant and interpretable in a biological context, reflecting the true nature of biological systems. To make simulations computationally tractable for large systems, CG models are often used and the impacts of the environmental conditions, such as ions and ligands, are not explicitly considered. Such simplifications may introduce biases or errors that affect the biological relevance of the simulation results, especially considering the complexity of large biological systems. Finally, the efficient storage and retrieval of extensive simulation data for public use remain formidable challenges amid the emergence of exascale simulations.

Looking to the future, the integration of technologies such as quantum computing with exascale systems may further revolutionize the field of biomolecular simulation. The future also holds the promise of increasingly sophisticated AI and deep learning models that can learn from and interpret the massive datasets generated by exascale simulations. In an exascale community, cloud computing will offer scalable and on-demand access to unprecedented computational speeds needed for biomedical computing [63].

Specific to nucleic acid simulations, exascale computing power in long-time, large-scale atomistic simulations could further significantly enhance our ability to predict RNA folding in vivo, including cotranscriptionally folding [64], to integrate experimental information into MD simulations for nucleic acid systems [65], to design and simulate RNA/DNA origami [66] and RNA riboswitches [67]. Furthermore, the optimization of classical force fields and the development of machine learning potentials for nucleic acids fall behind proteins. Historically, there has been a greater volume of research and development focused on proteins compared to nucleic acids. With the advent of exascale computing power, nucleic acid research has promising potential to advance at a rapid pace, especially for the large-scale structures such as ribosomes [68, 69], nucleosomes, chromatin, and chromosomes [70], and even the entirety of a cell [71].

Acknowledgments

This work was supported by the National Institutes of Health under Grant R35-GM134919 to S-J. C.

References

[1]. Gupta C, Sarkar D, Tieleman DP, Singharoy A, The ugly, bad, and good stories of large-scale biomolecular simulations, Curr Opin Struct Biol 73 (2022) 102338. [PubMed: 35245737]

[2]. Galvanetto N, Ivanović MT, Chowdhury A, Sottini A, Nüesch MF, Nettels D, Best RB, Schuler B, Extreme dynamics in a biomolecular condensate, Nature 619 (7971) (2023) 876–883. [PubMed: 37468629]

[3]. Wieczor M, Genna V, Aranda J, Badia RM, Gelpí JL, Gapsys V, de Groot BL, Lindahl E, Municoy M, Hospital A, et al. , Pre-exascale HPC approaches for molecular dynamics simulations. COVID-19 research: A use case, Wiley Interdiscip Rev Comput Mol Sci 13 (1) (2023) e1622. *

The study highlights the pivotal role of pre-exascale HPC in enhancing MD simulations, critical for advancing COVID-19 research.

[4]. Melo MC, Bernardi RC, Fostering discoveries in the era of exascale computing: How the next generation of supercomputers empowers computational and experimental biophysics alike, Biophys J 122 (14) (2023) 2833–2840. [PubMed: 36738105]

[5]. Chang C, Deringer VL, Katti KS, Van Speybroeck V, Wolverton CM, Simulations in the era of exascale computing, Nat Rev Mater 8 (5) (2023) 309–313. [PubMed: 37168499]

[6]. Sinha S, Li J, Tam B, Wang SM, Classification of PTEN missense VUS through exascale simulations, Brief Bioinform 24 (6) (2023) bbad361. [PubMed: 37843401]

[7]. Shaw DE, Adams PJ, Azaria A, Bank JA, Batson B, Bell A, Bergdorf M, Bhatt J, Butts JA, Correia T, et al., Anton 3: twenty microseconds of molecular dynamics simulation before lunch, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021, pp. 1–11.

[8]. Shim KS, Greskamp B, Towles B, Edwards B, Grossman J, Shaw DE, The specialized high-performance network on Anton 3, in: 2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA), IEEE, 2022, pp. 1211–1223. * Anton 3 introduces a high-performance network that significantly enhances parallel MD simulations by optimizing computational resource usage.

[9]. Andoh Y, Ichikawa S.-i., Sakashita T, Fujimoto K, Yoshii N, Nagai T, Tang Z, Okazaki S, An exa-scale high-performance molecular dynamics simulation program: MODYLAS, J Chem Phys 158 (19) (2023) 194803. [PubMed: 37184018]

[10]. Sheng N, Tong Z, Jiang C, Ma X, Yang X, Li H, Tian Z, Zhang Q, Microsecond simulation in a special-purpose molecular dynamics computer cluster, in: 2023 11th International Conference on Bioinformatics and Computational Biology (ICBCB), IEEE, 2023, pp. 151–157.

[11]. Voelz VA, Pande VS, Bowman GR, Folding@home: achievements from over twenty years of citizen science herald the exascale era, Biophys J 122 (14) (2023) 2852–2863. [PubMed: 36945779] * Folding@home's distributed computing has advanced biomolecular simulation and contributed to COVID-19 antiviral research.

[12]. Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, Meller A, Mallimadugula UL, Kuhn CE, Borowsky JH, Wiewiora RP, et al. , SARS-CoV-2 simulations go exascale to capture spike opening and reveal cryptic pockets across the proteome, BioRxiv (2020) 2020–06.

[13]. Raghavan B, Paulikat M, Ahmad K, Callea L, Rizzi A, Ippoliti E, Mandelli D, Bonati L, De Vivo M, Carloni P, Drug design in the exascale era: A perspective from massively parallel QM/MM simulations, J Chem Inf Model 63 (12) (2023) 3647–3658. [PubMed: 37319347]

[14]. Liebl K, Zacharias M, The development of nucleic acids force fields: From an unchallenged past to a competitive future, Biophys J 122 (14) (2022) 2841–2851. [PubMed: 36540025] * This review discusses the evolution of nucleic acids force fields, acknowledging their limitations and the emergence of improved models.

[15]. Winkler L, Galindo-Murillo R, Cheatham TE III, Structures and dynamics of DNA mini-dumbbells are force field dependent, J Chem Theory Comput 19 (8) (2023) 2198–2212. [PubMed: 36976268]

[16]. Tan D, Piana S, Dirks RM, Shaw DE, RNA force field with accuracy comparable to state-of-the-art protein force fields, Proc Natl Acad Sci USA 115 (7) (2018) E1346–E1355. [PubMed: 29378935]

[17]. Tucker MR, Piana S, Tan D, LeVine MV, Shaw DE, Development of force field parameters for the simulation of single-and double-stranded dna molecules and DNA-protein complexes, J Phys Chem B 126 (24) (2022) 4442–4457. [PubMed: 35694853]

[18]. Liebl K, Zacharias M, Tumuc1: a new accurate DNA force field consistent with high-level quantum chemistry, J Chem Theory Comput 17 (11) (2021) 7096–7105. [PubMed: 34662102]

[19]. Mlynsky V, Kuhrova P, Stadlbauer P, Krepl M, Otyepka M, Banas P, Sponer J, Simple adjustment of intranucleotide base-phosphate interaction in the OL3 AMBER force field improves RNA simulations, J Chem Theory Comput 19 (22) (2023) 8423–8433. [PubMed: 37944118]

[20]. Winkler L, Cheatham TE III, Benchmarking the Drude polarizable force field using the r(GACC) tetranucleotide, J Chem Inf Model 63 (8) (2023) 2505–2511. [PubMed: 36996447]

[21]. Winkler L, Galindo-Murillo R, Cheatham TE III, Assessment of A- to B-DNA transitions utilizing the drude polarizable force field, J Chem Theory Comput 19 (23) (2023) 8955–8966. [PubMed: 38014857]

[22]. George A, Mondal S, Purnaprajna M, Athri P, Review of electrostatic force calculation methods and their acceleration in molecular dynamics packages using graphics processors, ACS Omega 7 (37) (2022) 32877–32896. [PubMed: 36157750]

[23]. Chollet I, Lagardère L, Piquemal J-P, ANKH: A generalized O(N) interpolated Ewald strategy for molecular dynamics simulations, J Chem Theory Comput 19 (10) (2023) 2887–2905. [PubMed: 37134146] ** ANKH offers a linear-complexity, parallel-optimized method for calculating electrostatic energy in large-scale MD simulations.

[24]. Chen H, Chipot C, Enhancing sampling with free-energy calculations, Curr Opin Struct Biol 77 (2022) 102497. [PubMed: 36410221]

[25]. Prajapati JD, Onuchic JN, Sanbonmatsu KY, Exploring the energy landscape of riboswitches using collective variables based on tertiary contacts, J Mol Biol 434 (18) (2022) 167788. [PubMed: 35963460]

[26]. Caino-Lores S, Cuendet M, Marquez J, Kots E, Estrada T, Deelman E, Weinstein H, Taufer M, Runtime steering of molecular dynamics simulations through in situ analysis and annotation of collective variables, in: Proceedings of the Platform for Advanced Scientific Computing Conference, 2023, pp. 1–11.

[27]. Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, Rother KM, Bujnicki JM, SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction, Nucleic Acids Res 44 (7) (2016) e63–e63. [PubMed: 26687716]

[28]. Li J, Zhang J, Wang J, Li W, Wang W, Structure prediction of RNA loops with a probabilistic approach, PLoS Comput Biol 12 (8) (2016) e1005032. [PubMed: 27494763]

[29]. Zhang D, Li J, Chen S-J, IsRNA1: de novo prediction and blind screening of RNA 3D structures, J Chem Theory Comput 17 (3) (2021) 1842–1857. [PubMed: 33560836]

[30]. Li J, Chen S-J, RNAJP: enhanced RNA 3D structure predictions with non-canonical interactions and global topology sampling, Nucleic Acids Res 51 (7) (2023) 3341–3356. [PubMed: 36864729]

[31]. Mu Z-C, Tan Y-L, Zhang B-G, Liu J, Shi Y-Z, Ab initio predictions for 3D structure and stability of single-and double-stranded DNAs in ion solutions, PLoS Comput Biol 18 (10) (2022) e1010501. [PubMed: 36260618]

[32]. Cragnolini T, Laurin Y, Derreumaux P, Pasquali S, Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings, J Chem Theory Comput 11 (7) (2015) 3510–3522. [PubMed: 26575783]

[33]. Šulc P, Romano F, Ouldridge TE, Doye JP, Louis AA, A nucleotide-level coarse-grained model of RNA, J Chem Phys 140 (23) (2014) 235102. [PubMed: 24952569]

[34]. Šulc P, Romano F, Ouldridge TE, Rovigatti L, Doye JP, Louis AA, Sequence-dependent thermodynamics of a coarse-grained DNA model, J Chem Phys 137 (13) (2012) 135101. [PubMed: 23039613]

[35]. Li J, Chen S-J, RNA 3D structure prediction using coarse-grained models, Front Mol Biosci 8 (2021) 720937. [PubMed: 34277713]

[36]. Mu Z-C, Tan Y-L, Liu J, Zhang B-G, Shi Y-Z, Computational modeling of DNA 3D structures: From dynamics and mechanics to folding, Molecules 28 (12) (2023) 4833. [PubMed: 37375388]

[37]. Li Z, Portillo-Ledesma S, Schlick T, Brownian dynamics simulations of mesoscale chromatin fibers, Biophys J 122 (14) (2023) 2884–2897. [PubMed: 36116007]

[38]. Wasim A, Bera P, Mondal J, Development of a data-driven integrative model of a bacterial chromosome, J Chem Theory Comput 20 (4) (2024) 1673–1688. [PubMed: 37083406]

[39]. Lappala A, Wang CY, Kriz A, Michalk H, Tan K, Lee JT, Sanbonmatsu KY, Four-dimensional chromosome reconstruction elucidates the spatiotemporal reorganization of the mammalian X chromosome, Proc Natl Acad Sci USA 118 (42) (2021) e2107092118. [PubMed: 34645712]

[40]. Tan C, Jung J, Kobayashi C, Torre DUL, Takada S, Sugita Y, Implementation of residue-level coarse-grained models in GENESIS for large-scale molecular dynamics simulations, PLoS Comput Biol 18 (4) (2022) e1009578. [PubMed: 35381009] * This study introduces

an integrative approach for coarse-graining modeling of large-scale proteins, RNA, and DNA structures.

[41]. Wong CK, Tang C, Schreck JS, Doye JP, Characterizing the free-energy landscapes of DNA origamis, Nanoscale 14 (7) (2022) 2638–2648. [PubMed: 35129570]

[42]. Poppleton E, Romero R, Mallya A, Rovigatti L, Šulc P, OxDNA.org: a public Webserver for coarse-grained simulations of DNA and RNA nanostructures, Nucleic Acids Res 49 (W1) (2021) W491–W498. [PubMed: 34009383]

[43]. Zhou Y, Jiang Y, Chen S-J, RNA-ligand molecular docking: Advances and challenges, Wiley Interdiscip Rev Comput Mol Sci 12 (3) (2022) e1571. [PubMed: 37293430]

[44]. Olenginski LT, Kasprzak WK, Attionu SK, Shapiro BA, Dayie TK, Virtual screening of hepatitis B virus pre-genomic RNA as a novel therapeutic target, Molecules 28 (4) (2023) 1803. [PubMed: 36838792]

[45]. Paternoga H, Crowe-McAuliffe C, Bock LV, Koller TO, Morici M, Beckert B, Myasnikov AG, Grubmüller H, Nováček J, Wilson DN, Structural conservation of antibiotic interaction with ribosomes, Nat Struct Mol Biol 30 (9) (2023) 1380–1392. [PubMed: 37550453]

[46]. Wang Y, Parmar S, Schneekloth JS, Tiwary P, Interrogating RNA–small molecule interactions with structure probing and artificial intelligence-augmented molecular simulations, ACS Cent Sci 8 (6) (2022) 741–748. [PubMed: 35756372]

[47]. Gapsys V, Hahn DF, Tresadern G, Mobley DL, Rampp M, de Groot BL, Pre-exascale computing of protein–ligand binding free energies with open source software for drug design, J. Chem. Inf. Model 62 (5) (2022) 1172–1177. [PubMed: 35191702]

[48]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. , Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589. [PubMed: 34265844]

[49]. Simpkin AJ, Mesdaghi S, Sánchez Rodríguez F, Elliott L, Murphy DL, Kryshtafovych A, Keegan RM, Rigden DJ, Tertiary structure assessment at CASP15, Proteins 91 (12) (2023) 1616–1635. [PubMed: 37746927]

[50]. Das R, Kretsch RC, Simpkin AJ, Mulvaney T, Pham P, Rangan R, Bu F, Keegan R, Topf M, Rigden D, et al. , Assessment of three-dimensional RNA structure prediction in CASP15, bioRxiv (2023) 2023–04.

[51]. Li J, Zhang S, Chen S-J, Advancing RNA 3D structure prediction: Exploring hierarchical and hybrid approaches in CASP15, Proteins 91 (12) (2023) 1779–1789. [PubMed: 37615235]

[52]. Chen S-J, Hassan M, Jernigan RL, Jia K, Kihara D, Kloczkowski A, Kotelnikov S, Kozakov D, Liang J, Liwo A, et al. , Protein folds vs. protein folding: Differing questions, different challenges, Proc Natl Acad Sci USA 120 (1) (2023) e2214423119. [PubMed: 36580595]

[53]. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Schütt KT, Tkatchenko A, Müller K-R, Machine learning force fields, Chem Rev 121 (16) (2021) 10142–10186. [PubMed: 33705118]

[54]. Bhatia H, Aydin F, Carpenter TS, Lightstone FC, Bremer P-T, Ingólfsson HI, V Nissley D, Streitz FH, The confluence of machine learning and multiscale simulations, Curr Opin Struct Biol 80 (2023) 102569. [PubMed: 36966691] * This study reviews the integration of machine learning into multiscale simulations.

[55]. Unke OT, Stöhr M, Ganscha S, Unterthiner T, Maennel H, Kashubin S, Ahlin D, Gastegger M, Sandonas LM, Tkatchenko A, et al. , Accurate machine learned quantum-mechanical force fields for biomolecular simulations, arXiv preprint arXiv:2205.08306 (2022).

[56]. Durumeric AE, Charron NE, Templeton C, Musil F, Bonneau K, Pasos-Trejo AS, Chen Y, Kelkar A, Noé F, Clementi C, Machine learned coarse-grained protein force-fields: Are we there yet?, Curr Opin Struct Biol 79 (2023) 102533. [PubMed: 36731338]

[57]. Li J, Zhu W, Wang J, Li W, Gong S, Zhang J, Wang W, RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks, PLoS Comput Biol 14 (11) (2018) e1006514. [PubMed: 30481171]

[58]. Townshend RJ, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, Dror RO, Geometric deep learning of RNA structure, Science 373 (6558) (2021) 1047–1051. [PubMed: 34446608]

[59]. Eastman P, Galvelis R, Peláez RP, Abreu CR, Farr SE, Gallicchio E, Gorenko A, Henry MM, Hu F, Huang J, et al. , OpenMM 8: Molecular dynamics simulation with machine learning potentials, arXiv preprint arXiv:2310.03121 (2023).

[60]. Do HN, Wang J, Miao Y, Deep learning dynamic allostery of G-protein-coupled receptors, bioRxiv (2023) 2023–01.

[61]. Martin W, Sheynkman G, Lightstone FC, Nussinov R, Cheng F, Interpretable artificial intelligence and exascale molecular dynamics simulations to reveal kinetics: Applications to Alzheimer's disease, Curr Opin Struct Biol 72 (2022) 103–113. [PubMed: 34628220]

[62]. Liao SE, Sudarshan M, Regev O, Deciphering RNA splicing logic with interpretable machine learning, Proc Natl Acad Sci USA 120 (41) (2023)e2221165120. [PubMed: 37796983] * This study introduces an interpretable machine learning model that predicts RNA splicing from sequences, thereby deepening our understanding of their interplay.

[63]. Kutzner C, Kniep C, Cherian A, Nordstrom L, Grubmüller H, de Groot BL, Gapsys V, GROMACS in the cloud: A global supercomputer to speed up alchemical drug design, J Chem Inf Model 62 (7) (2022) 1691–1711. [PubMed: 35353508]

[64]. Xu X, Jin L, Xie L, Chen S-J, Landscape zooming toward the prediction of RNA cotranscriptional folding, J Chem Theory Comput 18 (3) (2022) 2002–2015. [PubMed: 35133833]

[65]. Bernetti M, Bussi G, Integrating experimental data with molecular simulations to investigate RNA structural dynamics, Curr Opin Struct Biol 78 (2023) 102503. [PubMed: 36463773]

[66]. Poppleton E, Urbanek N, Chakraborty T, Griffo A, Monari L, Göpfrich K, RNA origami: design, simulation and application, RNA Biol 20(1) (2023) 510–524. [PubMed: 37498217]

[67]. Mukherjee S, Barash D, Designing RNA switches for synthetic biology using inverse-RNA-folding, Trends Biotechnol (2023).

[68]. Nishima W, Girodat D, Holm M, Rundlet EJ, Alejo JL, Fischer K, Blanchard SC, Sanbonmatsu KY, Hyper-swivel head domain motions are required for complete mRNA-tRNA translocation and ribosome resetting, Nucleic Acids Res 50 (14) (2022) 8302–8320. [PubMed: 35808938]

[69]. Girodat DJ, Wieden H-J, Blanchard SC, Sanbonmatsu KY, Geometric alignment of aminoacyl-tRNA relative to catalytic centers of the ribosome underpins accurate mRNA decoding, Nat Commun 14 (2023) 5582. [PubMed: 37696823]

[70]. Sanbonmatsu KY, Large-scale simulations of nucleoprotein complexes: ribosomes, nucleosomes, chromatin, chromosomes and CRISPR, Curr Opin Struct Biol 55 (2019) 104–113. [PubMed: 31125796]

[71]. Thornburg ZR, Bianchi DM, Brier TA, Gilbert BR, Earnest TM, Melo MC, Safronova N, Sáenz JP, Cook AT, Wise KS, et al. , Fundamental behaviors emerge from simulations of a living minimal cell, Cell 185 (2) (2022) 345–360. [PubMed: 35063075] ** A dynamic model of a minimal cell with a small genome reveals fundamental insights into the balance of cellular processes and the principles of minimal life.

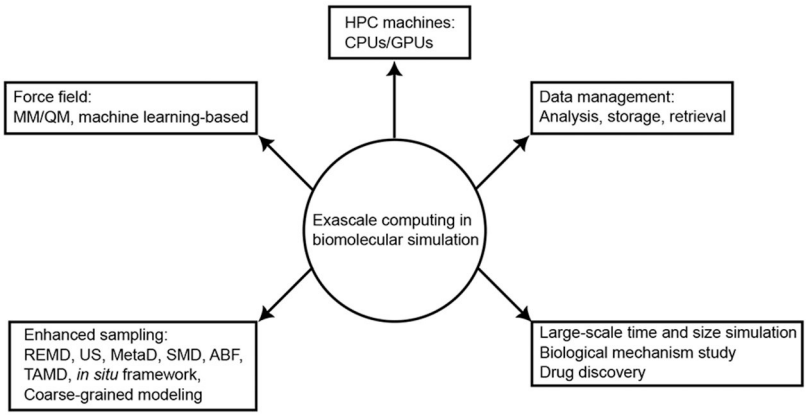

Figure 1:

Diagrammatic representation of core hardware and software elements in exascale computing systems for biomolecular simulations with potential applications.