Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Biomolecular simulations at the exascale: From drug design to organelles and beyond

Vytautas Gapsys[1], Wojciech Kopec[2,3], Dirk Matthes[3] and Bert L. de Groot[3]

The rapid advancement in computational power available for research offers to bring not only quantitative improvements, but also qualitative changes in the field of biomolecular simulation. Here, we review the state of biomolecular dynamics simulations at the threshold to exascale resources becoming available. Both developments in parallel and distributed computing will be discussed, providing a perspective on the state of the art of both. A main focus will be on obtaining binding and conformational free energies, with an outlook to macromolecular complexes and (sub)cellular assemblies.

Addresses
[1] Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30, Beerse 2340, Belgium
[2] Department of Chemistry, Queen Mary University of London, 327 Mile End Road, London E1 4NS, UK
[3] Computational Biomolecular Dynamics Group, Max Planck Institute for Multidisciplinary Sciences, Am Fassberg 11, 37077 Göttingen, Germany

Corresponding author: de Groot, Bert L. (bgroot@gwdg.de)
(Gapsys V.), (Kopec W.),

Current Opinion in Structural Biology 2024, **88**:102887

This review comes from a themed issue on **Theory and Simulation/ Computational Methods (2024): Biomolecular simulation in the Exascale Era and beyond**

Edited by **Paolo Carloni** and **Karissa Y. Sanbonmatsu**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2024.102887

Introduction

Molecular dynamics (MD) simulations have come of age [1]. Since their first application to biomolecular systems almost fifty years ago [2], both algorithmic developments as well as ever growing computational power have contributed to MD simulations now being an integral part of molecular biophysics. From deciphering protein function to computational drug design, from enzyme catalysis to the simulation of whole viruses or organelles, the field has matured substantially over these years. In the example of classical atomistic MD simulations, the state of the art currently reaches milliseconds for small proteins on one side of the spectrum to nanoseconds for hundreds of millions of atom supramolecular complexes on the other side [3,4] (Figure 1).

Either by use of massively parallel simulation strategies or being able to access longer timescales, ever better statistics become available for quantitative simulation analysis. Ranging from alchemical methods in protein or drug design, to conformational landscapes sampled using collective coordinates employing a variety of methods, and to stitching simulations from massively parallel trajectories together via Markov state modeling and then assessing state populations from the statistics, methods to assess free energies have become commonplace.

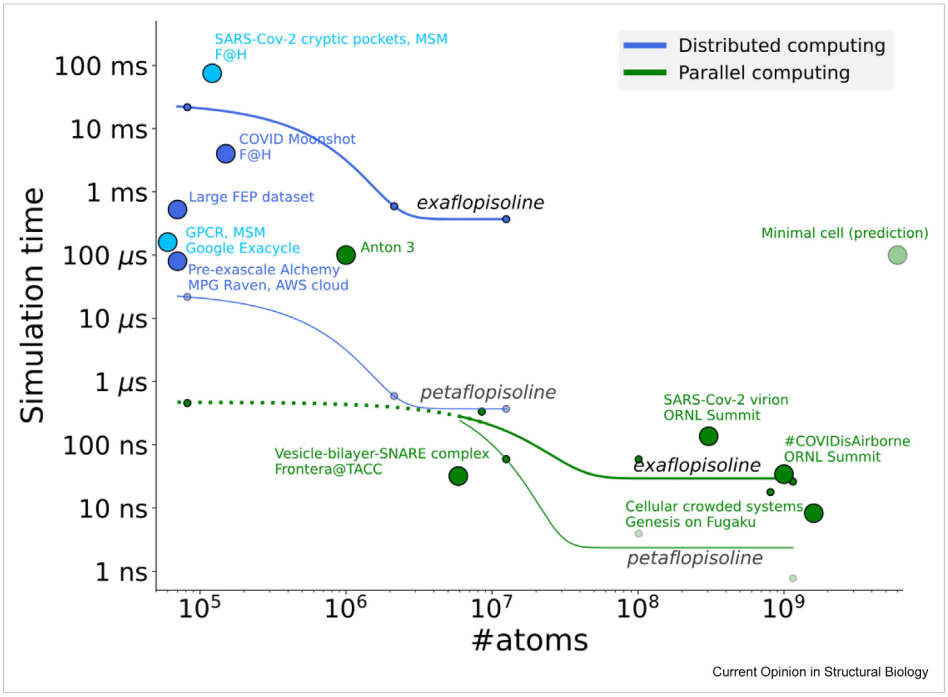
Currently at the dawn of exascale compute resources becoming available [5], opportunities are emerging across the field. In this review, we will concentrate on a subset of these developments with a focus on free energies, including alchemical free energy calculations of ligand–protein interactions, conformational free energy landscapes of individual biomolecules, as well as the sampling of supramolecular assemblies. We specifically focus on molecular dynamics simulations with molecular mechanics force fields, while we can recommend other reviews describing current state-of-the-art for large-scale biomolecular simulations with QM [6] and QM/MM [7].

Biomolecular simulations in drug design campaigns

In anticipation of the upcoming exascale resources, a number of studies reported on preparatory work to meet the challenge of efficiently utilizing such resources for alchemical free energy calculations. These studies on their own are still far from reaching the exascale level, rather they report on a concept of how protein-ligand binding free energy calculations can be distributed, deployed, and subsequently analyzed in an automated manner at a large scale.

In a proof of concept study, more than 1600 protein-ligand binding affinity differences were calculated by simultaneously utilizing 480 nodes on the MPG Garching supercomputer Raven [15]. Subsequently, these

Figure 1



Approximation of the simulation timescales for atomistic molecular dynamics simulations achievable on an exascale resource in 24 h of uninterrupted sampling. The upper blue line corresponds to the distributed computing scenario; the lower green line marks the true parallelization of a simulated system across an exascale machine. The isolines are based on the benchmarks in Refs. [8–10]: an exponential fit was used to extrapolate the results of the benchmarks to exa- and petaflop range (marked in small circles). Subsequently, another exponential fit was used to obtain the isolines through the extrapolated points. With the increase in system size, the achievable sampling time should eventually drop to zero, it is only due to the slow exponential decay the isolines appear to plateau for the larger system sizes. The dotted green line for smaller system sizes denotes the approximate range where parallelizing on peta- or exascale HPC machines is not efficient, i.e. using the full compute resource would rather degrade the performance. The simulation times were normalized to 4 fs integration time step. The large circles denote real simulation results without extrapolations. F@H SARS-Cov-2 cryptic pocket exploration [3] and the large FEP calculation [11] are marked assuming that all the calculations in the study could be performed in 24 h. Similarly, for COVID Moonshot [12] it was assumed that a single 'Sprint' would be performed in 24 h. Minimal cell system marked in the figure has not been simulated so far, but its size and aspired simulation time are described in Ref. [13]: we consider an atomistic representation of this system. References for the other marked studies: GPCR, MSM on Google Exacycle [14], Pre-exascale alchemy [15,10], Anton 3 [16], Vesicle-bilayer-SNARE complex [17], SARS-Cov-2 virion [9], #COVIDisAirborne [18], and Cellular crowded systems [8].

simulations were repeated on an AWS cloud infrastructure [10]. These studies, however, are only a preparatory step for approaching true exascale (Figure 1).

In one of the largest retrospective alchemical studies to date, Ross et al. [11] performed alchemical calculations for more than 1200 compounds (more than 2100 edges). This effort exceeds 500 μs of total sampling time. While such sampling is readily achievable with the currently available computational resources, access to an exascale machine would allow to run all the simulations for this study in less than a day (Figure 1). Such a throughput would considerably speed up prospective protein-ligand binding affinity predictions in contemporary drug design campaigns [19,20].

In a community effort to find SARS-Cov-2 main protease (MPro) inhibitors [12], the established consortium gathered experts in structural biology as well as medicinal and computation chemistry. The designed compounds were screened by means of alchemical relative binding free energy calculations prior to synthesis and evaluation in experimental assays. Simulations for the project were set up to make use of the Folding@home resources, which have recently reached an exascale mark [3]. The free energy calculations were performed in batches, termed 'Sprints'. Assuming that one Sprint could be performed in one day (although this does not necessarily need to be the case) and considering a typical cumulative timescale noted in Ref. [12], we see that this alchemical approach could indeed come close to the exascale threshold for distributed computing (Figure 1).

The distributed computing character of many drug design campaigns renders it suitable to scale to large numbers of compounds by massively parallel deployment on exascale machines or distributed compute networks. Challenges therefore are mostly anticipated in the preparatory/management stages, efficient process orchestration as well as for data analysis, which should

ideally be integrated in the workflows rather than post-processed.

Sampling protein conformations with distributed computing

The kinetics of many biological phenomena, such as protein-ligand or protein−protein interactions, important to protein function and drug development can require millisecond to second sampling time.

However, even with exascale machines, it is not expected that simulation timescales beyond milliseconds will be immediately accessible for larger biomolecular complexes, as classical MD simulations will ultimately be limited by the need for small integration steps to avoid numerical error accumulation [21,22]. Therefore, distributed computing is not only applicable to large-scale alchemical free energy calculations, but also holds the promise to provide the computational resources and sampling strategies necessary to simulate the substantial conformational changes of proteins that occur on biologically relevant timescales. Using the first-ever exascale computing platform, created by interconnected personal computers from the Folding@home community of volunteers, Bowman and co-workers were able to scrutinize the opening dynamics of the SARS-CoV-2 spike protein complex and predict cryptic binding pockets near the active site and dimerization interface of MPro [3]. In fact, projects like Folding@home [23,24] and GPUGrid [25] have paved the way for massively parallel approaches to all-atom biomolecular simulations, by making use of graphics-processing unit acceleration, enhanced sampling methodologies and statistical methods such as Markov State Models [26] in MD-based high-performance computing [14].

A huge number of short, completely independent, or loosely coupled trajectories of MD simulations run in conjunction with biasing schemes [22,27], can be used to efficiently explore simulation systems even with complex underlying free energy landscapes. In a recent example, simulations (on Folding@home) of the 299-amino acid long apolipoprotein E (ApoE) monomer made it possible to sample heterogenous protein conformations on the millisecond time scale and enabled model building to compare to and corroborate experimental single molecule studies of the system [28]. Structural ensembles generated on massively parallel computer architectures equivalent to millisecond MD simulations made it feasible to gain atomistic insight into the transiently populated, partially ordered states of the N-terminal domain of p53 [29], the binding of ligands to the intrinsically disordered protein p27 [30] and catalytically relevant conformational dynamics of the protein methyltransferase SETD8 [31].

Over the past two decades Anton supercomputers [16], a family of special-purpose hardware spearheaded by D.E. Shaw Research and designed to run long-timescale MD simulations have pushed Moore's limit for time-per-MD step. This notable, alternative approach to generic CPU/GPU machines reportedly reaches 100 μs per day, for a 1 million atom system in the latest generation, Anton 3 [16]. Although the direct FLOPs comparison between general and special-purpose machines is cumbersome, projecting the achievable sampling times in Figure 1 places Anton 3 well above the petaflop isoline. Recent accounts of MD simulations in the tens to hundreds of microsecond range achieved with Anton include the cancer drug imatinib binding to Abelson tyrosine kinase [32], the binding of small-molecule compound fasudil to monomeric α-synuclein [33] and the discovery of a mutant-selective inhibitor of the fibroblast growth factor receptor 2 [34].

Large systems requiring massive parallelization

The continuous progress toward exascale allows researchers to build and simulate molecular systems of increased complexity, in recent years particularly focusing on whole virus particles (virions), individual organelles [35], subcellular architectures, such as the nuclear pore complex [36] or an entire gene locus [37], or the bacterial cytoplasm [38,8]. Each of these systems presented an outstanding computational challenge, and, accordingly, provided unique insights, not attainable through simulations on a smaller scale. Indeed, the ever-increasing complexity and size of simulated biomolecular systems are important not only from the boundary-pushing perspective, but also promise discoveries of unknown phenomena that arise only at such scales and conditions, markedly different from typical *in vitro*-like conditions of routine simulations and experiments.

The interest in modeling complete envelopes of virus particles has recently been fueled by the COVID-19 pandemic. To this end, initial computational efforts from the Amaro group focusing on virus particles of various influenza variants, reaching ca. 160 million atoms simulated for up to 440 ns [39], paved the way to simulations of a full SARS-CoV-2 viral envelope (∼305 million atoms, 84 ns) [9] and peaked with simulations of a SARS-CoV-2 virion in the respiratory aerosol environment, resulting in one of the largest molecular systems simulated to date, of 1 billion atoms (simulated for 2.42 ns, with the reported performance of 2.55 PFLOPS (34.21 ns/day), using NAMD on the 4096 ORNL Summit nodes) [18].

These first-of-its-kind simulations of an aerosolized virus open completely new research avenues, such as atomistic views of the virus infection. A similar performance of ca.

12 ns/day has been also reported for a similarly sized system (>1 billion of atoms) representing bacterial cytoplasm, using GENESIS software on the FUGAKU supercomputer (and up to 8.30 ns for 1.62 billion atoms) [8].

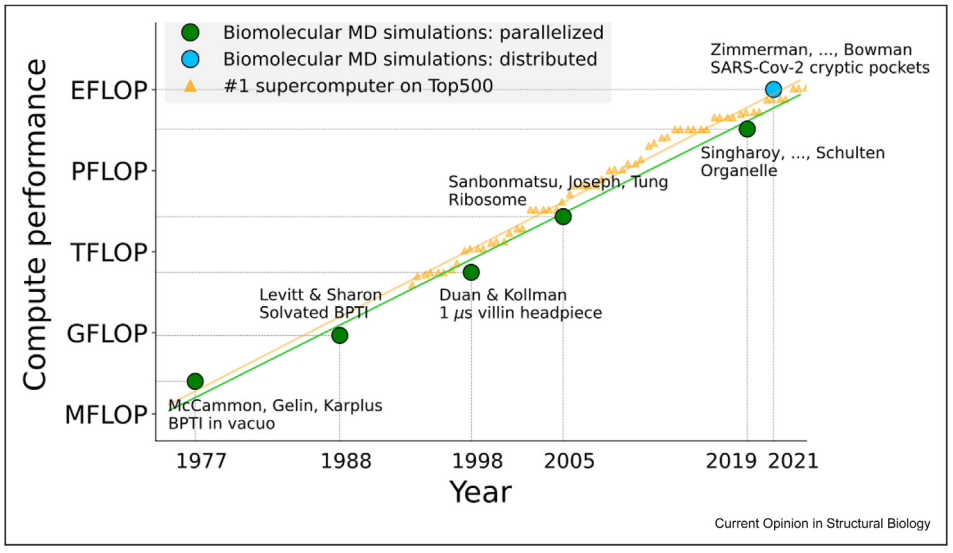
A natural next step in these modeling efforts is the construction and subsequent simulation of a whole-cell model. The availability of high resolution data of many cellular complexes, together with the integrative modeling tools, particularly AI-based, allowed researchers to build the first near-atomistic model of a whole minimal cell [13], Syn3A, owing to its extensive experimental characterization. The final system consisted of 550 million coarse-grained (MARTINI) particles (i.e. more than 6 billion atoms), involving a full chromosome, cytosol, and a lipid envelope. Putting aside the impressive feat of assembling such a massive system, it turned out the MD code of choice, GROMACS, was not ready (early 2023) to handle this system, thus the actual MD steps of this whole cell model have not been reported. The mentioned problems included handling multiple large molecules (e.g. the genome) spread over many domains in the domain decomposition approach, typically used to efficiently parallelize MD simulations. A related issue appeared in the simulations of the SARS-CoV-2 virion in the respiratory aerosol: a reduced particle density in an aerosol results in large 'empty' spaces, which is difficult to handle by standard domain decomposition. Finally, the amount of data generated by such simulations will be a challenge on its own.

These examples underlie the necessity of concomitant development of algorithms and novel software solutions (e.g. AI-driven), to be able to fully exploit the exascale (and beyond) resources for MD simulations. It has been nevertheless estimated that dedicated computational resources of the current type would enable reaching whole-cell simulations of 10–100 μs (taking into account the typical MARTINI timestep of 20–30 fs). The fully atomistic simulations of such a system, however, would require further advancements of molecular dynamics engines in terms of scaling. For example, the widespread Particle Mesh Ewald (PME) method used to handle long-range electrostatic interactions might become too inefficient in all-atom simulations of cell-like systems. More flexible approaches, such as Fast Multipole Method (FMM) [40] offer a beneficial alternative to PME.

Outlook

The widespread use of MD simulations and their applications to various scientific questions naturally led to very diverse approaches of executing them on distinct computational architectures in recent years. In this review, we provide an overview over recent state-of-the-art studies that approach the exascale computing range (Figure 1). Apart from millisecond timescales becoming more and more reachable, another emerging trend is moving away from single molecule systems towards biomolecular complexes, protein or ligand series, to organelles and (sub)cellular assemblies.

Figure 2



Historic overview of the hallmark biomolecular simulations. The green symbols mark largest systems that were amenable for molecular dynamics simulations in the respective year. The cyan symbol reaches the exascale threshold by employing distributed computing Folding@home platform to simulate multiple smaller systems. The orange symbols correspond to the performance of the most powerful supercomputer from the Top500 list [42]. References for the marked studies: BPTI in vacuo [2], Solvated BPTI [43], 1 μs villin headpiece [44], Ribosome [45], Organelle [35], and SARS-Cov-2 cryptic pockets [3].

Crossing the frontier to exascale computing power poses novel technical challenges concerning the administration and execution of the MD simulations (e.g. workflow management). These may eventually require building simulation architectures from "algorithms to transistors" [16] to make full use of capabilities provided by the (physical) exascale infrastructure, once its widespread availability is achieved. Preliminary studies have also already highlighted that the biosimulation community will likely have to evolve the ways by which the generated data is handled for analysis and shared (e.g. Exabyte problem) [41].

From a historical perspective, the compute resources used by biomolecular simulations follow closely the trend of the most powerful supercomputers (Figure 2). The extent, in terms of FLOPs, used by the hallmark applications of molecular dynamics grows exponentially, doubling every 15 months. At this rate, a 1000-fold increase in FLOP usage comes in about every 12 years. If the trend remains, the zettaflop barrier could be achieved in the next decade.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

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

* of special interest
** of outstanding interest

1. Herman JC Berendsen: **Bio-molecular dynamics comes of age**. *Science* 1996, **271**:954. 954.

2. Andrew McCammon J, Gelin Bruce R, Karplus Martin: **Dynamics of folded proteins**. *Nature* 1977, **267**:585–590.

3. ** Zimmerman Maxwell I, Porter Justin R, Ward Michael D, Singh Sukrit, Vithani Neha, Meller Artur, Mallimadugula Upasana L, Kuhn Catherine E, Borowsky Jonathan H, Wiewiora Rafal P, Hurley Matthew FD, Harbison Aoife M, Fogarty Carl A, Coffland Joseph E, Fadda Elisa, Voelz Vincent A, Chodera John D, Bowman Gregory R: **SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome**. *Nat Chem* 2021, **13**:651–659.
This study on the first exascale computing platform explores the conformational preferences of the SARS-CoV-2 spike protein and identifies potentially druggable cryptic pockets in the SARS-CoV-2 protein Nsp16.

4. Schade Robert, Kenter Tobias, Elgabarty Hossam, Lass Michael, Schütt Ole, Lazzaro Alfio, Pabst Hans, Mohr Stephan, Hutter Jürg, Kühne Thomas D, *et al.*: **Towards electronic structure-based ab-initio molecular dynamics simulations with hundreds of millions of atoms**. *Parallel Comput* 2022, **111**:102920.

5. Naddaf Miryam: **The science events to watch for in 2024**. *Nature* 2024, **625**:221–223.

6. Gavini Vikram, Baroni Stefano, Blum Volker, Bowler David R, Buccheri Alexander, Chelikowsky James R, Das Sambit, Dawson William, Delugas Pietro, Dogan Mehmet, *et al.*: **Roadmap on electronic structure codes in the exascale era**. *Model Simulat Mater Sci Eng* 2023, **31**, 063301.

7. Raghavan Bharath, Paulikat Mirko, Ahmad Katya, Callea Lara, Rizzi Andrea, Ippoliti Emiliano, Mandelli Davide, Bonati Laura, De Vivo Marco, Carloni Paolo: **Drug design in the exascale era: a perspective from massively parallel qm/mm simulations**. *J Chem Inf Model* 2023, **63**:3647–3658.

8. Jung Jaewoon, Kobayashi Chigusa, Kasahara Kento, Tan Cheng, Kuroda Akiyoshi, Minami Kazuo, Ishiduki Shigeru, Nishiki Tatsuo, Inoue Hikaru, Ishikawa Yutaka, *et al.*: **New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems**. *J Comput Chem* 2021, **42**:231–241.

9. Casalino Lorenzo, Dommer Abigail C, Gaieb Zied, Barros Emilia P, Sztain Terra, Ahn Surl-Hee, Trifan Anda, Brace Alexander, Bogetti Anthony T, Clyde Austin, *et al.*: **AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics**. *Int J High Perform Comput Appl* 2021, **35**:432–451.

10. * Kutzner Carsten, Kniep Christian, Cherian Austin, Nordstrom Ludvig, Grubmüller Helmut, Groot Bert L de, Gapsys Vytautas: **Gromacs in the cloud: a global supercomputer to speed up alchemical drug design**. *J Chem Inf Model* 2022, **62**:1691–1711.
Costs and efficiency assessment of state-of-the-art cloud computing and traditional on-premises computer clusters using the open-source software suite GROMACS for high-performance alchemical molecular dynamics of protein-ligand binding.

11. * Ross Gregory A, Lu Chao, Scarabelli Guido, Albanese Steven K, Houang Evelyne, Abel Robert, Harder Edward D, Wang Lingle: **The maximal and current accuracy of rigorous protein-ligand binding free energy calculations**. *Commun Chem* 2023, **6**:222.
In-depth comparison of differences between experimental benchmark datasets and relative protein–ligand binding affinities obtained through theoretical state-of-the-art free energy perturbation workflows.

12. ** Boby Melissa L, Fearon Daren, Ferla Matteo, Filep Mihajlo, Koekemoer Lizbé, Robinson Matthew C, Moonshot Consortium‡ COVID, Chodera John D, Lee Alpha A, Nir London, *et al.*: **Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors**. *Science* 2023, **382**, eabo7201.
Crowdsourced exascale molecular simulations utilized together with machine learning and high-throughput structural biology enable drug design of noncovalent and nonpeptidic main protease inhibitor scaffold.

13. Stevens Jan A, Grünewald Fabian, Marco van Tilburg PA, König Melanie, Gilbert Benjamin R, Brier Troy A, Thornburg Zane R, Luthey-Schulten Zaida, Marrink Siewert J: **Molecular dynamics simulation of an entire cell**. *Front Chem* 2023, **11**, 1106495.

14. Kohlhoff Kai J, Shukla Diwakar, Lawrenz Morgan, Bowman Gregory R, Konerding David E, Belov Dan, Altman Russ B, Pande Vijay S: **Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways**. *Nat Chem* December 2013, **6**:15–21.

15. Gapsys Vytautas, Hahn David F, Tresadern Gary, Mobley David L, Rampp Markus, de Groot Bert L: **Pre-exascale computing of protein–ligand binding free energies with open source software for drug design**. *J Chem Inf Model* 2022, **62**:1172–1177.

16. Shaw David E, Adams Peter J, Azaria Asaph, Bank Joseph A, Batson Brannon, Bell Alistair, Bergdorf Michael, Bhatt Jhanvi, Adam Butts J, Correia Timothy, *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*; 2021:1–11.

17. Rizo Josep, Sari Levent, Qi Yife, Im Wonpil, Lin Milo M: **All-atom molecular dynamics simulations of Synaptotagmin-SNARE-complexin complexes bridging a vesicle and a flat lipid bilayer**. *Elife* jun 2022, **11**:e76356.

18. Dommer Abigail, Casalino Lorenzo, Kearns Fiona, Rosenfeld Mia, Wauer Nicholas, Ahn Surl-Hee, Russo John, Oliveira Sofia, Morris Clare, Bogetti Anthony, *et al.*: **#COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol**. *Int J High Perform Comput Appl* 2023, **37**:28–44.

19. Schindler Christina EM, Baumann Hannah, Blum Andreas, Dietrich Böse, Buchstaller Hans-Peter, Burgdorf Lars, Cappel Daniel, Chekler Eugene, Paul Czodrowski, Dorsch Dieter, *et al.*: **Large-scale assessment of binding free energy calculations in active drug discovery projects**. *J Chem Inf Model* 2020, **60**:5457–5474.

20. Özen Ayşegül, Perola Emanuele, Brooijmans Natasja, Kim Joseph: *Prospective application of free energy methods in drug discovery programs, chapter 5*. 2021:127–141.

21. Ossyra John, Sedova Ada, Tharrington Arnold, Noé Frank, Clementi Cecilia, Smith Jeremy C: **Porting adaptive ensemble molecular dynamics workflows to the summit supercomputer**. In *High performance computing*. Edited by Weiland Michèle, Juckeland Guido, Alam Sadaf, Jagode Heike, Cham: Springer International Publishing; 2019:397–417.

22. Zimmerman Maxwell I, Porter Justin R, Sun Xianqiang, Silva Roseane R, Bowman Gregory R: **Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the apparent mechanism of conformational changes**. *J Chem Theor Comput* 2018, **14**:5459–5475.

23. Shirts Michael, Pande Vijay S: **Screen savers of the world unite**. *Science* December 2000, **290**:1903–1904.

24. Voelz Vincent A, Pande Vijay S, Bowman Gregory R: **Folding@home: achievements from over 20 years of citizen science herald the exascale era**. *Biophys J* July 2023, **122**:2852–2863.

25. Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G: **High-throughput all-atom molecular dynamics simulations using distributed computing**. *J Chem Inf Model* March 2010, **50**:397–403.

26. Bowman Gregory R, Pande Vijay S, Noé Frank: *An introduction to Markov state models and their application to long timescale molecular simulation*. Netherlands, Dordrecht: Springer; 2014:1–6.

27. Pérez Adrià, Herrera-Nieto Pablo, Doerr Stefan, Fabritiis Gianni De: **AdaptiveBandit: a multi-armed bandit framework for adaptive sampling in molecular simulations**. *J Chem Theor Comput* 2020, **16**:4685–4693.

28. * Stuchell-Brereton Melissa D, Zimmerman Maxwell I, Miller Justin J, Mallimadugula Upasana L, Incicco J Jeremías, Roy Debjit, Smith Louis G, Cubuk Jasmine, Baban Berevan, DeKoster Gregory T, *et al.*: **Apolipoprotein E4 has extensive conformational heterogeneity in lipid-free and lipid-bound forms**. *Proc Natl Acad Sci USA* 2023, **120**, e2215371120.
The diverse structure ensemble adopted by the pathogenic apolipoprotein E4, implicated in Alzheimer's Disease as genetic risk factor, is scrutinized by atomistic millisecond simulations on Folding@home.

29. Herrera-Nieto Pablo, Pérez Adrià, De Fabritiis Gianni: **Characterization of partially ordered states in the intrinsically disordered n-terminal domain of p53 using millisecond molecular dynamics simulations**. *Sci Rep* 2020, **10**:12402.

30. Herrera-Nieto Pablo, Pérez Adrià, De Fabritiis Gianni: **Small molecule modulation of intrinsically disordered proteins using molecular dynamics simulations**. *J Chem Inf Model* 2020, **60**:5003–5010.

31. Chen Shi, Wiewiora Rafal P, Meng Fanwang, Babault Nicolas, Ma Anqi, Yu Wenyu, Qian Kun, Hu Hao, Zou Hua, Wang Junyi, *et al.*: **The dynamic conformational landscape of the protein methyltransferase setd8**. *Elife* 2019, **8**:e45403.

32. * Ayaz Pelin, Lyczek Agatha, Paung YiTing, Mingione Victoria R, Iacob Roxana E, de Waal Parker W, Engen John R, Seeliger Markus A, Shan Yibing, Shaw David E: **Structural mechanism of a drug-binding process involving a large conformational change of the protein target**. *Nat Commun* 2023, **14**:1885.
Extensive and unbiased, multi-μs MD simulations run on a special purpose machine reveal a local structural instability in the C-terminal lobe of Abl kinase during binding of the cancer drug imatinib.

33. Robustelli Paul, Ibanez-de Opakua Alain, Campbell-Bezat Cecily, Giordanetto Fabrizio, Becker Stefan, Zweckstetter Markus, Pan Albert C, Shaw David E: **Molecular basis of small-molecule binding to α-synuclein**. *J Am Chem Soc* 2022, **144**:2501–2510.

34. Subbiah Vivek, Sahai Vaibhav, Maglic Dejan, Bruderek Kamil, Touré B Barry, Zhao Songping, Valverde Roberto, O'Hearn Patrick J, Moustakas Demetri T, Schönherr Heike, *et al.*: **RLY-4008, the first highly selective FGFR2 inhibitor with activity across FGFR2 alterations and resistance mutations**. *Cancer Discov* 2023, **13**:2012–2031.

35. Singharoy Abhishek, Maffeo Christopher, Delgado-Magnero Karelia H, Swainsbury David JK, Sener Melih, Ulrich Kleinekathöfer, Vant John W, Nguyen Jonathan, Hitchcock Andrew, Isralewitz Barry, *et al.*: **Atoms to phenotypes: molecular design principles of cellular energy metabolism**. *Cell* 2019, **179**:1098–1111.

36. Mosalaganti Shyamal, Obarska-Kosinska Agnieszka, Siggel Marc, Taniguchi Reiya, Turoňová Beata, Zimmerli Christian E, Buczak Katarzyna, Schmidt Florian H, Margiotta Erica, Mackmull Marie-Therese, *et al.*: **AI-based structure prediction empowers integrative structural analysis of human nuclear pores**. *Science* 2022, **376**, eabm9506.

37. Jung Jaewoon, Nishima Wataru, Daniels Marcus, Bascom Gavin, Kobayashi Chigusa, Adedoyin Adetokunbo, Wall Michael, Lappala Anna, Phillips Dominic, Fischer William, *et al.*: **Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations**. *J Comput Chem* 2019, **40**:1919–1930.

38. Yu Isseki, Mori Takaharu, Ando Tadashi, Harada Ryuhei, Jung Jaewoon, Sugita Yuji, Feig Michael: **Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm**. *Elife* 2016, **5**:e19274.

39. * Casalino Lorenzo, Seitz Christian, Lederhofer Julia, Tsybovsky Yaroslav, Wilson Ian A, Kanekiyo Masaru, Amaro Rommie E: **Breathing and tilting: Mesoscale simulations illuminate influenza glycoprotein vulnerabilities**. *ACS Cent Sci* 2022, **8**:1646–1663.
The conformational dynamics of the main antigen targets, hemagglutinin and neuraminidase, are characterized in a realistic crowded environment on the surface of influenza A virions by fully atomistic MD simulations.

40. Kohnke Bartosz, Ullmann Thomas R, Beckmann Andreas, Kabadshow Ivo, Haensel David, Morgenstern Laura, Dobrev Plamen, Groenhof Gerrit, Kutzner Carsten, Hess Berk, *et al.*: **Gromex: a scalable and versatile fast multipole method for biomolecular simulation**. In *Software for exascale computing-SPPEXA 2016-2019*. Springer International Publishing; 2020:517–543.

41. Wieczór Miłosz, Genna Vito, Aranda Juan, Badia Rosa M, Lluís Gelpí Josep, Gapsys Vytautas, de Groot Bert L, Lindahl Erik, Municoy Martí, Hospital Adam, Orozco Modesto: **Pre-exascale hpc approaches for molecular dynamics simulations. covid-19 research: a use case**. *WIREs Computational Molecular Science* 2023, **13**:e1622.

42. Top500, howpublished = https://www.top500.org/statistics/perfdevel/, note = Accessed: 2023-December-20.

43. Levitt Michael, Sharon Ruth: **Accurate simulation of protein dynamics in solution**. *Proc Natl Acad Sci USA* 1988, **85**:7557–7561.

44. Duan Yong, Kollman Peter A: **Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution**. *Science* 1998, **282**:740–744.

45. Sanbonmatsu Kevin Y, Simpson Joseph, Tung Chang-Shung: **Simulating movement of trna into the ribosome during decoding**. *Proc Natl Acad Sci USA* 2005, **102**:15854–15859.