All-Atom Biomolecular Simulation in the Exascale Era[†]

Thomas L. Beck,[*,‡] Paolo Carloni,[¶,§] and Dilipkumar N. Asthagiri[‡]

‡*National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA*

¶*INM-9/IAS-5 Computational Biomedicine, Forschungszentrum Jülich, Wilhelm-Johnen-Straße, D-54245 Jülich, Germany*

§*Department of Physics, RWTH Aachen University, D-52078 Aachen, Germany*

E-mail: becktl@ornl.gov

Abstract

Exascale supercomputers have opened the door to dynamical simulations, facilitated by AI/ML techniques, that model biomolecular motions over unprecedented length and time scales. This new capability holds the potential to revolutionize our understanding of fundamental biological processes. Here we report on some of the major advances that were discussed at a recent CECAM workshop in Pisa, Italy, on the topic, with a primary focus on atomic-level simulations. First, we highlight examples of current large-scale biomolecular simulations and the future possibilities enabled by crossing the exascale threshold. Next, we discuss challenges to be overcome to optimize the usage of these powerful resources. Finally, we close by listing several grand challenge problems that could be investigated with this new computer architecture.

Introduction

The advent of exascale super-computer facilities such as Frontier at the Oak Ridge National Laboratory USA (accessible to the global research community starting in April, 2023), Aurora at the Argonne National Laboratory USA, and JUPITER, coming this year in Jülich, Germany is providing vast new capabilities for tackling challenging societal problems. These include drug discovery,[1] personalized medicine,[2] biotechnology,[3] and imaging science.[4] Coupled with the exascale machines and advanced simulation methods, the simultaneous growth of the sciences of big data and artificial intelligence built upon machine learning (AI/ML) has further propelled the ability to discover informative patterns in massive simulation and experimental data sets.[5,6]

Regarding biophysical simulation and modeling, exascale computing enabled by GPU-acceleration and high-speed interconnects has opened the possibility of large-scale simulations of biomolecules and their (membrane-bound) assemblies that begin to approach the micron length scale.[7–10] These studies hold the potential to connect specific molecular features to mechanical and chemical function in the complex environments inside and outside cells.

Two of us (TB and PC) co-organized a CECAM workshop several months ago on "Biomolecular simulation and machine learning in the exascale era: first applications and perspectives" (https://www.cecam.org/workshop-details/1224) to survey this exciting topic, with a longer-term goal of facilitating the transition of biomolecular computing to a growing list of exascale facilities.

Highly diverse, complementary, and cutting-edge topics included conformational sampling,[11,12] protein thermodynamics[13] and kinetics,[14] drug discovery,[15,16] biomachines,[17] large macromolecular complexes such as the nuclear pore complex[7] and the ribosome,[18] enzyme catalysis by scalable QM/MM molecular dynamics algorithms,[19] AI/ML techniques for bioinformatics[20] and simulation,[21] and coarse graining approaches.[12] In addition, several speakers from Jülich, Oak Ridge, and elsewhere, rooted in computer science and software engineering, discussed the core computational challenges and opportunities related to exascale computing

in biology.[22,23]

Other recent reviews and perspectives[24–27] have discussed extensively the current status of biomolecular simulation and modeling, including coarse-grained methods.[27] Here, we focus on atomic-level simulations and both the enormous potential and resulting challenges created by the emergence of exascale computing.

Scaling up biomolecular simulations

The collection of talks at the workshop illustrates just how far the computational modeling of biomolecules has come in the last few decades. Also apparent is the wide diversity of the scientific themes. Major progress has been made in terms of both the resolution and fidelity of the models as well as in the length and time scales accessible. For instance, MD simulations of single membrane-bound proteins (at the atomic level) were state-of-the-art at the beginning of the century.[28] Such simulations are now routine on small to medium-sized computer clusters available at most research institutions.

Leadership class HPC offers a major leap in both the resolution of the simulations and in the system size and duration. For example, Frontier consists of 9,408 compute nodes, each of which can readily handle about 10^7 particles at a simulation production rate of about 3 ns/day. The length scale of a composite particle with this many atoms is on the order of 50 nm, which is in the size range of a typical virus.[9] Exploiting coarse-grained force fields[27,29]could expand the domain of applications eventually to small cells over biologically relevant time scales (Fig. 1). Gaining a deeper understanding of coupled biomolecular processes in the dense, crowded intracellular space may ultimately lead to paradigm shifts in drug design (including the design of drug combinations that can operate at multiple locations and points in time).

Figure 1 highlights several innovative studies from the last decade made possible by current HPC systems. Examples of an entire cell organelle (photosynthetic chromatophore

vesicle)[30] and the minimal cell[31] are aspirational studies with impressive initial results. We note the impressive progress in simply building systems as complex as the aerosolized virus,[8] the organelle,[30] and the minimal cell.[31] The minimal cell is currently a coarse-grained model, with the all atom system expected to be around 6×10^9 atoms. Simulating such a system to begin to explore basic cellular processes now appears within reach, but with significant constraints on the accessible time scales (see below). Indeed, an example of MD scaling up

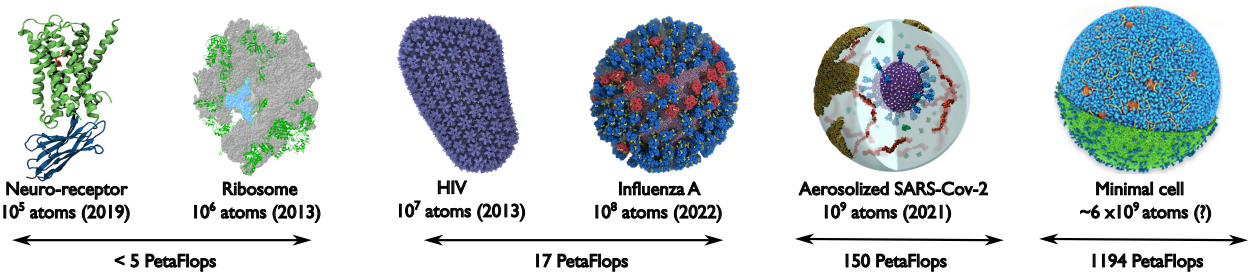

Figure 1: Select examples of innovative simulations over the last decade. From left to right: Neuro-receptor M_2 (adapted with permission from Figure 1 of Ref. 32; ©2019, American Chemical Society); Ribosome (PDB ID 4v6y, Ref. 33); HIV (PDB ID 3j3q, Ref. 34); Influenza A (image from www.eurekalert.org/multimedia/972194; credit Lorenzo Casalino, Amaro lab; original figure from Ref. 35, ©2022 American Chemical Society); Aerosolized SARS-Cov-2 (image from www.eurekalert.org/multimedia/900678; credit #COVIDisAirborne Team; original figure from Ref. 8, ©2021 Americal Chemical Society); and Minimal cell (adapted with permission from Figure 1 of Ref. 31; ©Stevens, Grünewald, van Tilburg, König, Gilbert, Brier, Thornburg, Luthey-Schulten and Marrink. The system comprises 561×10^6 beads, equivalent to about 6×10^9 atoms.) The HIV, Influenza, and SARS-Cov-2 simulations were performed using the leadership class HPC at ORNL. The compute speeds for the ORNL machines, including for Frontier, are based on the LINPACK benchmark. Ribosome and HIV images made using UCSF Chimera[36] and structures from the PDB.

to the full Frontier system is an exploratory simulation of bulk water comprising roughly 10^{11} atoms with a length scale approaching 1 micron and with long-range interactions fully involved (Hagerty and Asthagiri, unpublished). This length scale is typical for a micro-organism such as *E. Coli*. Assuming ideal weak scaling and thus again a simulation timescale production of 3 ns/day would imply consumption of ∼200,000 node-hours from a large allocation for a few-nanosecond simulation. These crude estimates provide a glimpse of the (distant) horizons that are now visible.

Of course many key cellular phenomena occur on a much longer time scale. While there

is merit in pushing the boundaries of molecular simulations into uncharted territory, it is important to keep in mind that interesting phenomena such as metabolism and replication are beyond what can be accessed. This has been discussed in detail in Ref. 26. Diffusion times in the viscous interior of cells, one of the simpler quantities that can be explored via simulation, grow quadratically with the distance (or as $\sim N^{2/3}$), so the movement of proteins and nucleic acids between organelles in cells is a major challenge for simulation.

Thus it is clear that simulating the dynamics of a large-scale system on a supercomputer for a short time, although indicating impressive progress, will not lead to insights into longer time scale phenomena (from large conformational changes to direct self-assembly that occur on the microsecond scales or longer[11]). Again, taking Frontier as an example, it is readily possible to simulate in less than an hour of real time a system with 10^6 atoms for 10^6 time steps using just $1/4^{\text{th}}$ of the resources per node (Asthagiri, unpublished). This suggests that a fruitful alternative approach is to examine an ensemble of smaller systems that can evolve over long time scales on each compute node, thus greatly enhancing the sampling and reducing uncertainties.[37] With $10^6 - 10^7$ atoms, however, the "smaller" systems above modeled by a single compute node are already capable of representing viruses.[38]

The tightly-coupled ensemble computing on a dedicated system like Frontier should be contrasted with highly distributed computing architectures (such as Folding@home[39]). The distributed computing paradigm of course has advantages of its own related to extensibility and optimal use of idle resources in order to conserve energy. The tightly-coupled nature of an exascale computer, however, allows for rapid-turnaround feedback both within the evolving simulation and in coupled interactions with ongoing experiments at remote facilities.

Through specialized ensemble techniques such as Markov State Models,[40] these computational developments create major opportunities for modelling the bio-assembly of large organized units. Further, many phenomena in biology, including enzyme catalysis, proton pumping, and electron transfer reactions, require a quantum mechanical treatment necessitating the use of QM/MM methods[19] or quantum-based surrogate models.[21] Some of them

have been massively parallelized in order to exploit exascale resources for modeling large systems.[41,42] Finally, while extensive work has gone into optimizing molecular force fields that mimic the basic interactions between atoms, it is now accepted that electronic polarization and many-body dispersion forces[43] can contribute substantially and can even qualitatively alter the equilibrium structures and self-assembly dynamics.[44] These effects will move to the fore and require additional capabilities as we attempt to model large assemblies at an accurate level.[45]

Integration of AI/ML in biomolecular simulations

A central theme gleaned from the workshop is the entry of AI/ML methods into nearly every domain of biomolecular simulation. The uses include AI-based structure prediction for large assemblies,[7] methods for increasing the accuracy and efficiency of the simulations via surrogate models,[21] coarse graining,[27] and tools for enhancing the sampling of conformational and other transitions between quasi-stable states.[11] Of course, AI/ML is also designed to find patterns in the massive datasets produced by the large-scale simulations.

The CECAM workshop suggested further ways in which AI/ML will accelerate discovery. As discussed above, a great benefit can be gained by dividing the modeling into a large ensemble that reduces statistical uncertainties and allows for the increased probability of sampling rare events. But on leadership-class machines, how to best organize these simulations is not a trivial task. In fact, the space of simulation run variables and parameters is very large, and a global optimization procedure driven by AI/ML is warranted. Utilization of AI/ML in this way could be viewed as "AI at the front end" or "computational design" (along the lines of "experimental design" in engineering systems, etc.[46]).

Further, methods are under development in which one can imagine high-level pruning or branching processes that select from emerging results to enhance sampling of the process of interest. This could be done as the simulations progress in order to minimize uncertainties

in the simulation outcome (see Ref. 47). This "on the fly" approach may accelerate discovery significantly. See also Ref. 8 for an example of such AI-guided sampling. Care must be taken in analyzing the influence of bias in the simulations from such enhanced sampling methods.

To summarize, a central suggestion of this perspective is a unified approach to the integration of modelling/simulation (ModSim) and AI/ML at all stages of the simulation process. At the front end, this involves building large biomolecular complexes that provide physically realistic starting models. Recently proposed Large Language Models (LLMs) will accelerate the process of generating alternative protein and RNA structures[48,49] and predicting functional protein sequences[50] in protein design. In addition, AI/ML methods can be used in a computational design process for the simulations in order to minimize uncertainties of the simulation results (above). For the simulations themselves, AI/ML methods (deep learning methods in particular) have been shown to greatly accelerate high-accuracy models generated from underlying quantum mechanical results via the generation of surrogate models such as DeePMD.[21] Finally, both modern experiments and large-scale simulations produce vast amounts of data that can be interrogated for patterns that shed light on the key motions occurring in the biological systems.

In addition, a feedback loop between the evolving simulations and uncertainty measures can be used to steer the simulations towards more robust and reliable results. The importance of uncertainty quantification methodologies is receiving extensive attention in the molecular simulation community.[37] Beyond this integration that accelerates simulation efficiency and convergence, AI/ML methods will also impact the nature of the experiment/theory interaction, discussed below.

Software engineering

Another conclusion from the workshop is the pressing need for greater software portability, data uniformity and ready access to the research community, and modern visualization and

analysis tools.[22,51] To some extent, the community, like several others, has developed at a rapid pace with many projects moving in parallel and often independently.

This is a natural progression, but, with the appearance of several exascale centers around the world, scientific output would be significantly enhanced by more coordinated collective efforts reminiscent of the Linux operating system development. Of particular focus should be the development of portable and efficient codes that can be implemented across a wide range of hardware offerings. Making data and workflows accessible to all users can aid in transparency, reproducibility, and could impact AI/ML methods broadly. A follow-on CECAM workshop on these topics will take place in May, 2024 (https://www.cecam.org/workshop-details/1319)

Integration with major experimental facilities

The current explosive growth in computational resources coincides with a rapid development of advanced light sources,[52] neutron scattering facilities,[53] (in cell) NMR,[54] and cryoEM and cryo-electron tomography techniques,[55,56] all of which generate large datasets. For example, the LCLS-II free electron X-ray laser facility at SLAC has recently come online, with a nearly 10,000 fold increase in pulse frequency compared to the existing LCLS beamline.[52,57] Exascale computing will play a key role in guiding and analyzing the resulting high fidelity experiments that hold the potential to reveal biomolecular structure and dynamics, producing "movies" of complex molecular-level processes in action.

Besides the enhanced pulse frequency of the LCLS-II X-ray free electron laser, an advantage is the ability to perform experiments at room temperature. Thus, realistic sampling of conformational states and binding events can be expected. A challenge will be dealing with the "data deluge" that could produce several petabytes of raw data per hour during an experiment.[58] The ability to handle massive datasets is already a feature of leadership-class (exascale) computing facilities. For example, the short term data storage capacity on the

Frontier system is 700 petabytes. A bottleneck is the current data transfer rate between facilities of about 400 gigabits/sec, necessitating some level of data reduction prior to transfer for real-time analysis.

These discussions foreshadow an evolution towards more tightly coupled experimental, theoretical, and simulation work that will accelerate the discovery process. One can imagine ongoing experiments at a light source or a cryoEM facility that produce preliminary results indicating a conformational transition between two functional states of a protein. Simultaneously, a simulation could be ongoing on the exascale super-computer that provides results related to the same possible transition. An indication of a third conformational state appears in the experiment, which provides a trigger to initiate further replica exchange simulations at higher temperatures. A structure similar to the third state seen in the experiment appears in the simulation, but the detailed agreement with experiment is poor. The experiment/theory comparison leads to the hypothesis that the charge state of a side-chain is in error in the computational model. Altering the charge state (or other components of the force field) on the fly then leads to better agreement with experiment.

Studies of this kind are currently possible through much slower step-by-step feedback that can take months or years rather than minutes or hours. We see this as the sign of a new, promising era for the interaction of experiment and theory, producing accelerated discovery for complex systems.

As sketched above, close coupling of powerful experimental and computational facilities in real time better leverages the ensemble computing capabilities of dedicated leadership computing facilities relative to distributed computing architectures such as Folding@home.[39] Indeed, efforts are currently underway within the US Department of Energy to develop the aforementioned tight integration within a larger science ecosystem.

Conclusions

We are at the threshold of a new era of biomolecular simulation. We can explore the driving forces for many essential processes performed by large assembled structures that are the workhorses in biological systems. Prototype systems of this kind include the nuclear pore complex,[15] the ribosome,[18,33] large membrane protein complexes,[59] microtubules,[60] and viruses.[8] Outside of the realm of molecular-level biophysical science, how these large complexes guide the coupled interactions in systems biology will be a further horizon that will begin to come into focus.

The capabilities that enable the above studies come with significant challenges related to optimizing and standardizing codes, computational workflows, data handling, and visualization. If these challenges are met, molecular dynamics simulations over biologically relevant timescales to probe key interactions that drive the functioning of organelles or even entire cells may soon be running on exascale machines. Progress in these directions can be expected to revolutionize our understanding of fundamental biological processes with applications that include drug discovery and biotechnology.

Acknowledgement

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. PC thanks the Helmholtz European Partnering Program ("Innovative high-performance computing approaches for molecular neuromedicine") for funding. The open access publication fee is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

References

(1) Song, L. F.; Merz Jr, K. M. Evolution of alchemical free energy methods in drug discovery. *J. Chem. Info. Mod.* **2020**, *60*, 5308–5318.

(2) Cirillo, D.; Valencia, A. Big data analytics for personalized medicine. *Curr. Op. Biotech.* **2019**, *58*, 161–167.

(3) Shajii, A.; Numanagić, I.; Leighton, A. T.; Greenyer, H.; Amarasinghe, S.; Berger, B. A Python-based programming language for high-performance computational genomics. *Nature Biotech.* **2021**, *39*, 1062–1064.

(4) Shamshad, F.; Khan, S.; Zamir, S. W.; Khan, M. H.; Hayat, M.; Khan, F. S.; Fu, H. Transformers in medical imaging: A survey. *Medical Image Analysis* **2023**, 102802.

(5) Agajanian, S.; Alshahrani, M.; Bai, F.; Tao, P.; Verkhivker, G. M. Exploring and learning the universe of protein allostery using artificial intelligence augmented biophysical and computational approaches. *J. Chem. Info. Mod.* **2023**, *63*, 1413–1428.

(6) Curioni, A. Artificial intelligence: Why we must get it right. *Informatik-Spektrum* **2018**, *41*, 7–14.

(7) Mosalaganti, S.; Obarska-Kosinska, A.; Siggel, M.; Taniguchi, R.; Turoňová, B.; Zimmerli, C. E.; Buczak, K.; Schmidt, F. H.; Margiotta, E.; Mackmull, M.-T., et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **2022**, *376*, eabm9506.

(8) Dommer, A.; Casalino, L.; Kearns, F.; Rosenfeld, M.; Wauer, N.; Ahn, S.-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A., et al. # COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol. *Intl. J. High Perf. Comput. Appl.* **2023**, *37*, 28–44.

(9) Goodsell, D. S.; Olson, A. J.; Forli, S. Art and science of the cellular mesoscale. *Trends Biochem. Sc.* **2020**, *45*, 472–483.

(10) Singharoy, A.; Pérez, A.; Chipot, C. Biophysics at the dawn of exascale computers. *Biophys. J.* **2023**, *122*, E1–E2.

(11) Jung, H.; Covino, R.; Arjun, A.; Leitold, C.; Dellago, C.; Bolhuis, P. G.; Hummer, G. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nature Comput. Sci.* **2023**, 1–12.

(12) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine learning for molecular simulation. *Ann. Rev. Phys. Chem.* **2020**, *71*, 361–390.

(13) Asthagiri, D. N.; Paulaitis, M. E.; Pratt, L. R. Thermodynamics of Hydration from the Perspective of the Molecular Quasichemical Theory of Solutions. *J. Phys. Chem. B* **2021**, *125*, 8294–8304.

(14) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Front. Mol. Biosci.* **2022**, *9*, 899805.

(15) Goßen, J.; Ribeiro, R. P.; Bier, D.; Neumaier, B.; Carloni, P.; Giorgetti, A.; Rossetti, G. AI-based identification of therapeutic agents targeting GPCRs: introducing ligand type classifiers and systems biology. *Chem. Sci.* **2023**, *14*, 8651–8661.

(16) Hsu, D. J.; Lu, H.; Kashi, A.; Matheson, M.; Gounley, J.; Wang, F.; Joubert, W.; Glaser, J. TwoFold: Highly accurate structure and affinity prediction for protein-ligand complexes from sequences. *Intl. J. High Perf. Comput. Appl.* **2023**, 10943420231201151.

(17) Dayananda, A.; Dennison, T.; Fonseka, H. Y. Y.; Avestan, M. S.; Wang, Q.; Tehver, R.; Stan, G. Allosteric communication in the gating mechanism for controlled protein degradation by the bacterial ClpP peptidase. *J. Chem. Phys.* **2023**, *158*.

(18) Ren, P.; Bhattacharjee, A.; Piana-Agostinetti, S.; Shaw, D. E. Toward long, all-atom simulation of the E. coli 70S ribosome. *Biophys. J.* **2022**, *121*, 204a.

(19) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Bircher, M. P.; Carloni, P.; Rothlisberger, U. MiMiC: a novel framework for multiscale modeling in computational chemistry. *J. Chem. Theory Comput.* **2019**, *15*, 3810–3823.

(20) Dias, R.; Evans, D.; Chen, S.-F.; Chen, K.-Y.; Loguercio, S.; Chan, L.; Torkamani, A. Rapid, Reference-Free human genotype imputation with denoising autoencoders. *Elife* **2022**, *11*, e75600.

(21) Wang, H.; Zhang, L.; Han, J.; Weinan, E. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184.

(22) Lanrezac, A.; Férey, N.; Baaden, M. Wielding the power of interactive molecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022**, *12*, e1594.

(23) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*.

(24) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J., et al. Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *Wiley Inter. Rev.: Comput. Molec. Sci.* **2019**, *9*, e1393.

(25) Schlick, T.; Portillo-Ledesma, S. Biomolecular modeling thrives in the age of technology. *Nature Comput. Sci.* **2021**, *1*, 321–331.

(26) Gupta, C.; Sarkar, D.; Tieleman, D. P.; Singharoy, A. The ugly, bad, and good stories of large-scale biomolecular simulations. *Curr. Op. Struct. Biol.* **2022**, *73*, 102338.

(27) Borges-Araújo, L.; Patmanidis, I.; Singh, A. P.; Santos, L. H.; Sieradzan, A. K.; Vanni, S.; Czaplewski, C.; Pantano, S.; Shinoda, W.; Monticelli, L., et al. Pragmatic Coarse-Graining of Proteins: Models and Applications. *J. Chem. Theory Comput.* **2023**,

(28) Bostick, D. L.; Berkowitz, M. L. Exterior site occupancy infers chloride-induced proton gating in a prokaryotic homolog of the ClC chloride channel. *Biophys. J.* **2004**, *87*, 1686–1696.

(29) Jin, J.; Pak, A. J.; Durumeric, A. E.; Loose, T. D.; Voth, G. A. Bottom-up coarse-graining: Principles and perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759–5791.

(30) Singharoy, A.; Maffeo, C.; Delgado-Magnero, K. H.; Swainsbury, D. J.; Sener, M.; Kleinekathöfer, U.; Vant, J. W.; Nguyen, J.; Hitchcock, A.; Isralewitz, B., et al. Atoms to phenotypes: molecular design principles of cellular energy metabolism. *Cell* **2019**, *179*, 1098–1111.

(31) Stevens, J. A.; Grünewald, F.; van Tilburg, P. M.; König, M.; Gilbert, B. R.; Brier, T. A.; Thornburg, Z. R.; Luthey-Schulten, Z.; Marrink, S. J. Molecular dynamics simulation of an entire cell. *Front. Chem.* **2023**, *11*, 1106495.

(32) Capelli, R.; Bochicchio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 3354–3361.

(33) Bock, L. V.; Blau, C.; Schröder, G. F.; Davydov, I. I.; Fischer, N.; Stark, H.; Rodnina, M. V.; Vaiana, A. C.; Grubmüller, H. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature Struct. Molec. Biol.* **2013**, *20*, 1390–1396.

(34) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497*, 643–646.

(35) Casalino, L.; Seitz, C.; Lederhofer, J.; Tsybovsky, Y.; Wilson, I. A.; Kanekiyo, M.; Amaro, R. E. Breathing and Tilting: Mesoscale Simulations Illuminate Influenza Glycoprotein Vulnerabilities. *ACS Cent. Sc.* **2022**, *8*, 1646–1663.

(36) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(37) Wan, S.; Sinclair, R. C.; Coveney, P. V. Uncertainty quantification in classical molecular dynamics. *Phil. Trans. Roy. Soc. A* **2021**, *379*, 20200082.

(38) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. Molecular dynamics simulations of the complete Satellite Tobacco Mosaic Virus. *Structure* **2006**, *14*, 437–449.

(39) Voelz, V. A.; Pande, V. S.; Bowman, G. R. Folding@ home: achievements from over twenty years of citizen science herald the exascale era. *Biophys. J.* **2023**,

(40) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.

(41) 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. Info. Mod.* **2023**,

(42) Manathunga, M.; Aktulga, H. M.; Götz, A. W.; Merz, K. M. Quantum mechan-

ics/molecular mechanics simulations on NVIDIA and AMD graphics processing units. *J. Chem. Info. Mod.* **2023**, *63*, 711–717.

(43) Stöhr, M.; Tkatchenko, A. Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions. *Science Advances* **2019**, *5*, eaax0024.

(44) Jing, Z.; Rackers, J. A.; Pratt, L. R.; Liu, C.; Rempe, S. B.; Ren, P. Thermodynamics of ion binding and occupancy in potassium channels. *Chem. Sci.* **2021**, *12*, 8920–8930.

(45) Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature Chem.* **2017**, *9*, 1005–1011.

(46) Blanchard, A.; Sapsis, T. Output-weighted optimal sampling for Bayesian experimental design and uncertainty quantification. *SIAM/ASA J. Uncert. Quant.* **2021**, *9*, 564–592.

(47) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suar, E.; Zuckerman, D. M.; Chong, L. T. WESTPA: An interoperable, highly scalable software package for Weighted Ensemble simulation and analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.

(48) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130.

(49) Zhang, Y.; Lang, M.; Jiang, J.; Gao, Z.; Xu, F.; Litfin, T.; Chen, K.; Singh, J.; Huang, X.; Song, G., et al. Multiple sequence-alignment-based RNA language model and its application to structural inference. *bioRxiv* **2023**, 2023–03.

(50) Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos Jr, J. L.; Xiong, C.; Sun, Z. Z.; Socher, R., et al. Large language models generate functional protein sequences across diverse families. *Nature Biotech.* **2023**, 1–8.

(51) Martinez, X.; Krone, M.; Alharbi, N.; Rose, A. S.; Laramee, R. S.; O'Donoghue, S.; Baaden, M.; Chavent, M. Molecular graphics: bridging structural biologists and computer scientists. *Structure* **2019**, *27*, 1617–1623.

(52) Zielinski, K. A. et al. RNA structures and dynamics with Å resolution revealed by x-ray free-electron lasers. *Science Advances* **2023**, *9*, eadj3509.

(53) Ashkar, R.; Bilheux, H.; Bordallo, H.; Briber, R.; Callaway, D. J.; Cheng, X.; Chu, X.-Q.; Curtis, J. E.; Dadmun, M.; Fenimore, P., et al. Neutron scattering in the biological sciences: Progress and prospects. *Acta Cryst. D Struct. Biol.* **2018**, *74*, 1129–1168.

(54) Luchinat, E.; Banci, L. In-cell NMR: From target structure and dynamics to drug screening. *Curr. Op. Struct. Biol.* **2022**, *74*, 102374.

(55) Giri, N.; Roy, R. S.; Cheng, J. Deep learning for reconstructing protein structures from cryo-EM density maps: Recent advances and future directions. *Curr. Op. Struct. Biol.* **2023**, *79*, 102536.

(56) Young, L. N.; Villa, E. Bringing Structure to Cell Biology with Cryo-Electron Tomography. *Ann. Rev. Biophys.* **2023**, *52*, 573–595.

(57) Ishigami, I.; Sierra, R. G.; Su, Z.; Peck, A.; Wang, C.; Poitevin, F.; Lisova, S.; Hayes, B.; Moss, F. R.; Boutet, S.; Sublett, R. E.; Yoon, C. H.; Yeh, S.-R.; Rousseau, D. L. Structural insights into functional properties of the oxidized form of cytochrome c oxidase. *Nature Commun.* **2023**, *14*, 5752.

(58) Thayer, J. B.; Carini, G.; Kroeger, W.; O'Grady, C.; Perazzo, A.; Shankar, M.; Weaver, M. Building a data system for lcls-ii. 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). 2017; pp 1–4.

(59) Chatzigoulas, A.; Cournia, Z. Predicting protein–membrane interfaces of peripheral

membrane proteins using ensemble machine learning. *Briefings in Bioinformatics* **2022**, *23*, bbab518.

(60) Macke, A. C.; Kelly, M. S.; Varikoti, R. A.; Mullen, S.; Groves, D.; Forbes, C.; Dima, R. I. Microtubule severing enzymes oligomerization and allostery: a tale of two domains. *J. Phys. Chem. B* **2022**, *126*, 10569–10586.

Graphical TOC Entry