

Fostering discoveries in the era of exascale computing: how the next-generation of supercomputers empowers computational and experimental biophysics alike.

Marcelo C. R. Melo¹ and Rafael C. Bernardi^{1,*}

¹Auburn University, Department of Physics, Auburn University, Auburn, AL 36849

*Correspondence: rbernardi@auburn.edu

ABSTRACT Over a century ago, physicists started broadly relying on theoretical models to guide new experiments. Soon thereafter, chemists began doing the same. Now, biological research enters a new era when experiment and theory walk hand in hand. Novel software and specialized hardware became essential to understand experimental data and propose new models. In fact, current petascale computing resources already allow researchers to reach unprecedented levels of simulation throughput to connect *in silico* and *in vitro* experiments. The reduction in cost and improved access allowed a large number of research groups to adopt supercomputing resources and techniques. Here, we outline how large-scale computing has evolved to expand decades-old research, spark new research efforts, and continuously connect simulation and observation. For instance, multiple publicly and privately funded groups have dedicated extensive resources to develop artificial intelligence tools for computational biophysics, from accelerating quantum chemistry calculations to proposing protein structure models. Moreover, advances in computer hardware have accelerated data processing from single-molecule experimental observations, and simulations of chemical reactions occurring throughout entire cells. The combination of software and hardware have opened the way for exascale computing and the production of the first public exascale supercomputer, Frontier, inaugurated by the Oak Ridge National Laboratory in 2022. Ultimately, the popularization and development of computational techniques and the training of researchers to use them will only accelerate the diversification of tools and learning resources for future generations.

SIGNIFICANCE From the integration of artificial intelligence (AI) tools to the development of specialized hardware, computational biophysics has evolved to encompass state-of-the-art technologies in every stage of its scientific research efforts. The development of new computational tools had implications across all fields, leading public funding agencies to support the creation and use of supercomputers for biological research, and private sector investments to support numerous *in silico* drug development start ups. The next-generation of exascale supercomputers are already creating new opportunities for integrating computational modeling and experiments, pushing computational biophysics beyond explaining experimental results, fostering discoveries with an unprecedented level of detail.

INTRODUCTION

The history of Biophysics is paved by cross-disciplinary innovation. From Erwin Schrödinger's take on genetics using an information transfer perspective (1) to Hodgkin and Huxley's description of electric currents in cells (2), biology and mathematics, physiology and physics, have continuously worked together. With the advent of computers, new opportunities emerged for combining disciplines, as exemplified by the 2013 Nobel Prize awarded ten years ago to Karplus, Levitt, and Warshel for their contributions to theoretical chemistry. Software and hardware have become essential to many aspects of biophysical research, not only to make sense of experimental data but to propose and validate new models.

In this perspective, we will outline how large-scale computing has provided novel insights to classic research efforts and can continue to open new roads for the exploration of

biophysical phenomena. While we attempt to cover important computational techniques, detailing their many differences would be outside the scope of this Perspective. To the readers that are interested in details of the computational techniques, we suggest recent reviews in quantum-mechanics/molecular-mechanics, coarse-grained simulations, Brownian dynamics, among other techniques that benefit from exascale computing (3–8). Here, we will focus on the interplay between methodological advances and biological breakthroughs.

FROM MOLECULES TO CELLS

Protein structure prediction, one of the oldest biochemistry challenges tackled by computational biophysics, was also the focus of recent broad spread media attention due to the performance of Artificial Intelligence (AI)-based algo-

rhythms. The notion that a primary sequence of amino acid residues contained all the necessary information to determine a protein's three-dimensional structure captivated researchers for decades. Many attempted to use computers to impose physical and chemical laws onto unsuspecting bits representing atoms and bonds, and despite documented examples of sequence-similar, structure-dissimilar proteins in public databases (9), homology-based protein structure prediction enjoyed continuous successes over the years, with prominent examples being Modeller (10), Rosetta (11), and SWISS-MODEL (12). The new AI-era models such as AlphaFold (13), RoseTTAFold (14), and more recently, OmegaFold (15), also rely heavily on previously determined primary and tertiary structures of known proteins, either explicitly using sequence alignments or implicitly storing that information on millions of neural network parameters. They have now reached the point where almost 200 million structures, including the entire human proteome, have been computationally predicted (16). The unifying thread in this diverse set of approaches is the extensive use of state-of-the-art computer hardware and software to optimize models, execute algorithms, and analyze results. Specifically, the development of new graphic processing units (GPUs) and tensor processing units (TPUs) over the past decade, along with specialized machine learning packages such as Keras, TensorFlow, PyTorch, among others, now provides unprecedented training potential for new neural network models and applications.

Beyond predicting structures, the study of protein dynamics and properties has benefited from large-scale computing in many ways. Dedicated hardware built exclusively for molecular dynamics (MD) simulations has allowed unprecedented insight into small protein's folding pathways (17). With the advance of GPUs and TPUs, training neural networks has become faster and easier, opening new doors for academic research. For example, AI models trained on protein sequences to produce learned representations can now be used to predict features such as fold stability and mutation impact on protein function (18). The same infrastructure powers efforts to integrate AI into analysis of protein dynamics (19), such as the dimensionality reduction of an MD trajectory using autoencoders (20). These and other developments were only possible due to the creation of advanced neural network architectures, including case-specific autoencoders and recurrent neural networks. This type of novel insight into protein sequence and dynamics can significantly help AI-based drug development pipelines (21) and computer-aided protein design and evolution efforts (22).

A particularly interesting research endeavour where computational power has bridged the gap between theory and experiments is single-molecule force spectroscopy (SMFS), where unbinding or unfolding forces are used as probes for molecular interactions and energy landscapes (25, 26). The computational study of mechanostable molecular interactions started decades ago. The streptavidin:biotin complex, for example, widely used in experimental research, was simu-

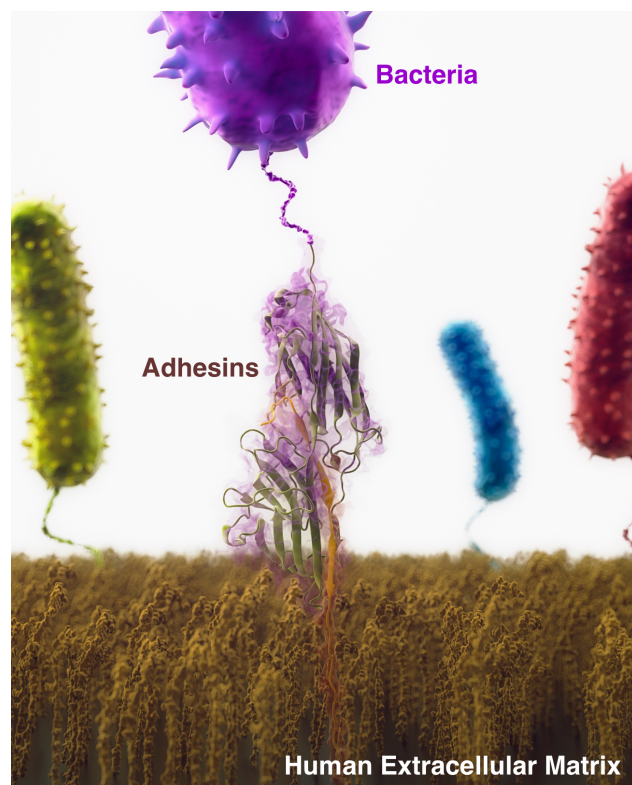


Figure 1: Modern supercomputers are powering discoveries with an *in silico* SMFS approach. *Staphylococcus aureus* is a pathogen that can form biofilms on implants and medical devices. A critical step to biofilm formation is the formation of a tight interaction between microbial surface proteins called adhesins and components of the extracellular matrix of the host. A combination of *in silico* and *in vitro* single-molecule techniques has revealed how the bond between staphylococcal adhesins and their human targets can withstand forces in the same order of that of covalent bonds (23, 24).

lated 25 years ago (27) using the most advanced hardware available at the time to achieve picosecond-level sampling of eight replicas per system. Almost twenty years later, the same molecular complex was revisited to elucidate the origins of conflicting experimental observations: In 2018 (28), a monovalent streptavidin tetramer was simulated to reveal different unbinding pathways for biotin when pulling the complex from C- or N-terminus; In 2020 (29), in a study that demanded 400 80-ns long replicas for a total of 32 μ s of simulation time, the four-subunit streptavidin complex bound to a single biotin molecule was simulated to reveal how the tethering geometry affects experimental measurements. Exploring more complicated interfaces, *in silico* SMFS reached hundreds of replicas totaling dozens of μ s of MD simulated time when modeling protein complexes, providing an unprecedented level of detail and explaining the origins of mechanostability in protein:protein interactions (23, 30–32). More recently, combining this *in silico* SMFS approach with AI-based structure

prediction has allowed for the first screening of mechanical properties of dozens of *Staphylococcus*' adhesins during initial steps of infection (24) (see Fig. 1). These efforts were made possible largely by the use of National Science Foundation (NSF)-funded supercomputers, which allowed parallel broad sampling of *in silico* SMFS experiments at multiple pulling speeds, tethering geometries, and interface variants. Despite technological advances, there is still a large gap between time scales accessible through experimental and *in silico* SMFS. While state-of-the-art resources have significantly broadened researcher's access to *all-atom* MD simulations that span millisecond time scales, few groups can reliably obtain enough replicas of independent simulations as to fit the models used to explain experimental observations. A recent example of this limitation was explored by using *coarse-grained* MD simulations to connect all-atom and experimental SMFS data for *Staphylococcus*' adhesins (33).

In recent decades, supercomputers have been used to simulate increasingly complicated molecular complexes (34). The first all-atom simulation of the entire HIV-1 viral capsid (35), for example, relied on cryo-electron-microscopy (Cryo-EM) data to build a 64 million atom system. In 2013, simulating this system required 128 thousand cores of the NSF-funded Blue Waters supercomputer to achieve 200ns of total simulation time for the entire capsid. Almost ten years later, the SARS-CoV-2 viral envelope was simulated using a combination of AI-powered methods and high-performance computing (HPC) resources from the Department of Energy-funded Summit supercomputer. This research led to the creation of a 305 million atom simulation that elucidated how elements of the SARS-CoV-2 viral envelope interact (36), which still achieved 68 nanoseconds-per-day of simulation performance. Large-scale simulation efforts have also targeted realistic compositions of cell membranes (37), and the electron transfer process in mitochondrial complexes (38).

More than broadening the reach of simulations to increasingly larger systems, high-performance computing has also been used to extend the range of enhanced sampling and free energy calculation in MD simulations (39). Many macromolecular complexes would require *seconds* of simulated time in order to spontaneously visit relevant biological states multiple times. That is why the continuous improvement of enhanced sampling methods is crucial for computational biophysics. For example, researchers used a weighted ensemble strategy to simulate the spike protein trimer of SARS-CoV-2 in over 130 μ s of trajectories, obtaining hundreds of spontaneous opening and closing events (40) that showed how glycans control the motion of this macromolecular complex. One particular challenge for the exascale era that is already being addressed today is the integration of quantum-mechanical/molecular-mechanical (QM/MM) calculations to classical simulations, which will require efficient communication as both simulation size and computer clusters grow over time (41, 42). QM/MM simulations have been used to investigate not only chemical reactions in biological molecules (43), but also to

understand how these molecules behave in highly polarizable environments, such as biological membranes (44). In a hybrid QM/MM approach, the charging of a tRNA molecule was examined using a massively parallel computation strategy that included a string method with swarms of trajectories to enhance the sampling of possible chemical reactions paths (45).

While exascale computing achieved by large supercomputer clusters will revolutionize computational biophysics research in the coming years, petascale research is already accessible today. The combination of drastic increase in computing power within similar price ranges for CPUs, the development of new architectures for GPUs, and improvements in supercomputer infrastructure and usability allows researchers to reach unprecedented levels of simulation throughput to connect *in silico* and *in vitro* experiments.

Improvements made to speed up MD software packages (46, 47), coupled with novel hardware, allowed researchers to accumulate collections of simulation data of unprecedented size. In fact, the development of specialized GPU kernels and the accelerated rate of inter GPU communication have greatly accelerated MD packages, both in small and large systems. Now, a major part of computational biophysics efforts are focused on the development of analysis methods that can accelerate data processing and extraction of information from simulations. Researchers are engaging this analysis challenge on many fronts, from graphical user interface (GUI)-based packages like QwikMD (48, 49), to coding-oriented packages such as MDAnalysis (50), PyTraj (51, 52), MDtraj (53), and Bio3D (54). Specialized analysis methods have also been developed to detect large-scale low frequency movements in macromolecular complexes (55, 56), or to represent biomolecules as a graph and detect cooperative motions (57, 58). The common theme among these and other methods is their ability to represent macromolecular dynamics in low-dimensional spaces that are amenable to both visualization and analysis. Thankfully, the origin of this problem also provides solutions. Computer clusters are regularly used to accelerate the analysis of results they created, and can even help visualizing systems too large to leave their massive storage devices (59, 60).

Experimental biology advances have been equally important in pushing the boundaries of computational biophysics research. Cryo-EM has occupied a prominent role in such studies given its ability to provide large-scale architectures for macromolecular complexes, and even detailed structures for their building blocks. From the structure of the eukaryotic translational initiation complex (61) to the structure of the mitochondrial complex I (62, 63) or the organization of the entire nuclear pore complex of *Saccharomyces cerevisiae* (64), Cryo-EM has continued to provide a macroscopic view for many research efforts, opening the doors for the detailed examination of dynamics that all-atom simulations can provide. A notable example was the determination of the water-mediated proton transport through the V_0 proton channel of a yeast ATPase (65), which revealed how the transport mechanism

relies on specific protonation states of amino acid side chains throughout the channel. The CryoEM and MD techniques outlined here can be combined for even more synergistic research, providing hybrid approaches for structure determination (66, 67), such as MD Flexible Fitting (68) and Damped-Dynamics Flexible Fitting (69). The exploration of macromolecular dynamics can also be accelerated by combining experiments and simulations(70), including methods such as CryoFold(71) and Metainference(72), where short-lived intermediate states can be computationally inferred to describe large-scale molecular rearrangements.

Moving towards cell-scale models, different simulation and analysis techniques are combined to reach length-scales and time-scales that would be impossible to achieve only a few years before. Already in 2008 researchers were exploring how crowding in the *Escherichia coli* cytoplasm affected molecular diffusion and bimolecular association reactions using coarse grained reaction diffusion simulations (73). In 2010, a landmark study used the same biological target, the *E. coli* cytoplasm, and its 50 most abundant macromolecules, to perform a Brownian dynamics (BD) simulation that assessed crowding effects on protein folding, molecular association, and thermodynamic properties in the cytoplasm (74). Recently, another group used an all-atom approach to build a model of the *E. coli* cytoplasm composed of 1.5 million atoms and submitted it to a total of 3 μ s of molecular dynamics simulation (75), focusing on the technical aspects of preparing and validating all-atom models for future larger-scale models of bacterial cytoplasm.

A *tour de force* that exemplifies the union of experimental and computational biophysics was the exploration of principles of cellular energy metabolism in *Rhodobacter sphaeroides* (76) (see Fig. 2). While all individual molecules and protein complexes involved in the photosynthetic energy conversion mechanism of the bacteria’s chromatophore were previously resolved, their structure and pairwise interactions are not sufficient to provide a unified model for this mechanism. It was necessary to combine atomic-force microscopy, electron microscopy, crystallography, mass spectrometry, proteomics, and optical spectroscopy data to build the model for the chromatophore. Beyond protein structures and the lipid composition of the membrane in which proteins were immersed, the very arrangement of all components was mapped to build a 100 million atom model for the organelle. With all this information, researchers could explore multiple aspects of the photosynthetic energy conversion mechanism. Simulations ranged from QM/MM, to obtain excitation energies of pigments, to all-atom, coarse grained, and BD, reaching hundreds of microseconds of sampling to track the internal movements of the electron carrier Cytochrome *c*₂. Evidently, the effort relied on supercomputing resources and a variety of specialized algorithms and software packages made available to the community.

Much of the work described so far focuses on studying a protein or complex within its local biological environment.

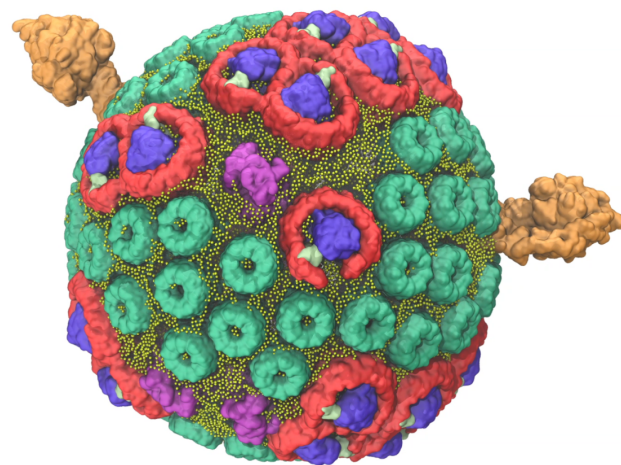


Figure 2: Model of an 100 million atom chromatophore, an organelle responsible for light-harvesting in *Rhodobacter sphaeroides*. The model was built using state-of-the-art supercomputers to combine atomic-force microscopy, electron microscopy, crystallography, mass spectrometry, proteomics, and optical spectroscopy data. Figure adapted from Singharoy et. al., 2019 (76).

Similarly, the study of chemical reactions can benefit from contextualizing them within the greater network to which they belong. The field of chemical kinetics, and in particular the discreteness and stochasticity inherent to chemical reactions within living cells, has seen great advances with the development of theory and computational methods for time-resolved simulations of reaction networks (77). Such simulations allow researchers to explore a variety of behaviors observed during cell growth, proliferation, and differentiation, for example with the modeling of self-regulated feedback loops in gene expression circuits (78).

While many reactions in a cell can be approximated using deterministic models, such as ordinary differential equations (ODE) or flux balance analysis, in order to reach whole-cell coverage one must integrate biochemical systems that rely on species with low copy numbers, such as mRNAs or transcription factors. In such cases, stochasticity is unavoidable, and researchers have developed software packages to describe reactions kinetics that can even incorporate spacial inhomogeneity. A prominent example was the recent whole-cell model of the model organism JCVI-syn3A, a minimal cell with an artificially reduced genome that allowed for unprecedented model coverage of a living organism (79). This effort counted with a variety of modeling approaches, and integrated ODEs with reaction diffusion master equation (RDME) models to describe an entire cell cycle of this cell, including energy metabolism, DNA replication, and gene expression. The same research also counted with an extensive array of experimental input, from qPCR to Cryo-electron tomograms.

In the context of whole-cell modeling, the development of

large-scale computing resources was doubly important. First, it allowed for rare events to be extensively sampled, which is essential for the study of stochastic systems and for validating models using experimental observations. Second, it allowed researchers to meet the parameter optimization demands from numeric approximations of complex reaction networks. Cell-scale models are inevitably underinformed, and researchers depend on parameter estimation and careful sensitivity analysis to properly analyze model results. StochSS (80), for example, is a stochastic simulation package already integrated into cloud computing environments, allowing researchers to access extensive computational power with ease. Lattice Microbes (81), on the other hand, focused on GPU-specialized code to extend the reach of RDME simulations into larger spacial coverage of a cell and longer simulation times, leveraging the impressive advances seen in GPU hardware over the past decade.

CONCLUSION

Since all software packages described here are freely available for academic use, and many are easily accessible through web interfaces, it is easy to forget about the costs associated with creating and providing such services. AlphaFold, for example, reportedly took many days of training using the most advanced hardware to reach the published model, and is distributed with over two terabytes of databases. Training the final version of this model (and ignoring intermediate versions) was estimated to require dozens of thousands of dollars in computing costs, resources largely unavailable to the vast majority of academic researchers in the world. Thinking about the big picture, the need for extensive computing resources in many fields justifies investments of hundreds of millions of dollars made by governments to build the fastest supercomputers in the world. Universities also recognize the need for computational infrastructure by funding new computing-oriented faculty with larger startup packages that approach the ones offered to experimental counterparts. This hidden cost of large-scale computing comes with an environmental impact: Australian astronomers estimated that their supercomputer usage is responsible for emitting the equivalent of approximately 15 kilotons of CO₂ per year (82). This is prompting supercomputing centers and cloud computing providers to be more transparent about energy expenditure, helping researchers to balance computational power and environmental impact (83). Hopefully, a combination of more efficient software and renewable energy sources will soon power all our research needs.

Between state-of-the-art HPC infrastructure easily accessible through cloud instances and breakthroughs in quantum computing already being used in drug development efforts (84), computational biophysics continues to benefit from private sector investments. NVIDIA routinely makes their hardware available for universities, so academic researchers can develop new algorithms using the latest advances in GPU

architectures, and AMD's GPUs currently power some of the largest supercomputers in the world. Microsoft, Amazon, and other cloud service providers offer free resources for students and academic rates for laboratories, not to mention the tech giants such as Google and Facebook who are continuously funding algorithmic developments in their AI-specialized divisions. In fact, both basic and applied research in computational biophysics now enjoy the neural network architectures originally designed for computer vision or text processing. While there are notable case-specific developments, such as AlphaFold, recurrent neural networks originally designed for language processing now power DNA sequence analysis, and facial recognition algorithms now process microscopy images.

While computational research in biophysics has always been driven by improvements in experimental methodologies, much of the past research has focused on establishing computational methods themselves. Now, with the advances made in computer hardware, *in silico* models can capitalize on the algorithms and theories developed over the past decades to literally reach the time- and length-scales observed experimentally. As a result, computational research can routinely and accurately create new proteins, propose atomic-level modifications to small molecules, and design regulatory circuits using molecular biology building blocks. More than simply recapitulating experimental observations, the current range of possibilities in computational biophysics allows us bring detail and provide missing information that is out of reach for *in vitro* models. Due to these and other advances, *in silico* research has gained both the attention and trust of experimentalists, which is reflected in major combined computational-experimental efforts now frequently seen in high-impact publications. Our path forward will include new combinations of computational and experimental efforts, with unprecedented precision and reach, ultimately allowing us to challenge biological assumptions and propose new models for biophysical phenomena.

AUTHOR CONTRIBUTIONS

All authors wrote and reviewed the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant MCB-2143787 (CAREER: In Silico Single-Molecule Force Spectroscopy).

REFERENCES

1. Schrodinger, E., 1951. What is life? The physical aspect of the living cell. At the University Press.
2. Hodgkin, A. L., and A. F. Huxley, 1952. A quantitative

- description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology* 117:500.
3. Hollingsworth, S. A., and R. O. Dror, 2018. Molecular dynamics simulation for all. *Neuron* 99:1129–1143.
 4. Venable, R. M., A. Kramer, and R. W. Pastor, 2019. Molecular dynamics simulations of membrane permeability. *Chemical reviews* 119:5954–5997.
 5. Salo-Ahen, O. M., I. Alanko, R. Bhadane, A. M. Bonvin, R. V. Honorato, S. Hossain, A. H. Juffer, A. Kabe-dev, M. Lahtela-Kakkonen, A. S. Larsen, et al., 2020. Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes* 9:71.
 6. Joshi, S. Y., and S. A. Deshmukh, 2021. A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation* 47:786–803.
 7. Huber, G. A., and J. A. McCammon, 2019. Brownian dynamics simulations of biological molecules. *Trends in chemistry* 1:727–738.
 8. Gupta, C., D. Sarkar, D. P. Tieleman, and A. Singharoy, 2022. The ugly, bad, and good stories of large-scale biomolecular simulations. *Current Opinion in Structural Biology* 73:102338.
 9. Kosloff, M., and R. Kolodny, 2008. Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins: Structure, Function, and Bioinformatics* 71:891–902.
 10. Webb, B., and A. Sali, 2016. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics* 54:5–6.
 11. Leman, J. K., B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, et al., 2020. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature methods* 17:665–680.
 12. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumieny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, et al., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46:W296–W303.
 13. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 1–11.
 14. Baek, M., F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, et al., 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373:871–876.
 15. Wu, R., F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, et al., 2022. High-resolution de novo structure prediction from primary sequence. *BioRxiv*.
 16. Callaway, E., 2022. ‘The entire protein universe’: AI predicts shape of nearly every known protein. *Nature*.
 17. Shaw, D. E., J. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, et al., 2014. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In SC’14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 41–53.
 18. Alley, E. C., G. Khimulya, S. Biswas, M. AlQuraishi, and G. M. Church, 2019. Unified rational protein engineering with sequence-based deep representation learning. *Nature methods* 16:1315–1322.
 19. Noé, F., G. De Fabritiis, and C. Clementi, 2020. Machine learning for protein folding and dynamics. *Current Opinion in Structural Biology* 60:77–84.
 20. Wehmeyer, C., and F. Noé, 2018. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of chemical physics* 148:241703.
 21. Melo, M. C., J. R. Maasch, and C. de la Fuente-Nunez, 2021. Accelerating antibiotic discovery through artificial intelligence. *Communications biology* 4:1–13.
 22. Wittmann, B. J., Y. Yue, and F. H. Arnold, 2021. Informed training set design enables efficient machine learning-assisted directed protein evolution. *Cell Systems* 1–20.
 23. Milles, L. F., K. Schulten, H. E. Gaub, and R. C. Bernardi, 2018. Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science* 359:1527–1533.
 24. Gomes, P. S. F. C., D. E. B. Gomes, and R. C. Bernardi, 2022. Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy. *Frontiers in Bioinformatics* 2.
 25. Žoldák, G., and M. Rief, 2013. Force as a single molecule probe of multidimensional protein energy landscapes. *Current opinion in structural biology* 23:48–57.
 26. Dudko, O. K., G. Hummer, and A. Szabo, 2006. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.* 96:108101.
 27. Izrailev, S., S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical journal* 72:1568–1581.

28. Sedlak, S. M., L. C. Schendel, M. C. Melo, D. A. Pippig, Z. Luthey-Schulten, H. E. Gaub, and R. C. Bernardi, 2018. Direction matters: Monovalent streptavidin/biotin complex under load. *Nano letters* 19:3415–3421.
29. Sedlak, S. M., L. C. Schendel, H. E. Gaub, and R. C. Bernardi, 2020. Streptavidin/biotin: Tethering geometry defines unbinding mechanics. *Science advances* 6:eay5999.
30. Bernardi, R. C., E. Durner, C. Schoeler, K. H. Malinowska, B. G. Carvalho, E. A. Bayer, Z. Luthey-Schulten, H. E. Gaub, and M. A. Nash, 2019. Mechanisms of nanonewton mechanostability in a protein complex revealed by molecular dynamics simulations and single-molecule force spectroscopy. *Journal of the American Chemical Society* 141:14752–14763.
31. Liu, Z., H. Liu, A. M. Vera, R. C. Bernardi, P. Tinnefeld, and M. A. Nash, 2020. High force catch bond mechanism of bacterial adhesion in the human gut. *Nature communications* 11:1–12.
32. Bauer, M. S., S. Gruber, A. Hausch, P. S. Gomes, L. F. Milles, T. Nicolaus, L. C. Schendel, P. L. Navajas, E. Procko, D. Lietha, et al., 2022. A tethered ligand assay to probe SARS-CoV-2:ACE2 interactions. *Proceedings of the National Academy of Sciences* 119:e2114397119.
33. Melo, M. C. R., D. E. B. Gomes, and R. C. Bernardi, 2023. Molecular Origins of Force-Dependent Protein Complex Stabilization during Bacterial Infections. *J. Am. Chem. Soc.* 145:70–77. <https://pubs.acs.org/doi/abs/10.1021/jacs.2c07674>.
34. Perilla, J. R., B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, 2015. Molecular dynamics simulations of large macromolecular complexes. *Current opinion in structural biology* 31:64–74.
35. Zhao, G., J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, et al., 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 497:643–646.
36. Casalino, L., A. C. Dommer, Z. Gaieb, E. P. Barros, T. Sztain, S.-H. Ahn, A. Trifan, A. Brace, A. T. Bogetti, A. Clyde, et al., 2021. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. *The International Journal of High Performance Computing Applications* 35:432–451.
37. Ingólfsson, H. I., M. N. Melo, F. J. Van Eerden, C. Arnarez, C. A. Lopez, T. A. Wassenaar, X. Periole, A. H. De Vries, D. P. Tieleman, and S. J. Marrink, 2014. Lipid organization of the plasma membrane. *Journal of the american chemical society* 136:14554–14559.
38. Galemou Yoga, E., K. Parey, A. Djurabekova, O. Haapanen, K. Siegmund, K. Zwicker, V. Sharma, V. Zickermann, and H. Angerer, 2020. Essential role of accessory subunit LYRM6 in the mechanism of mitochondrial complex I. *Nature communications* 11:1–8.
39. Bernardi, R. C., M. C. Melo, and K. Schulten, 2015. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1850:872–877.
40. Sztain, T., S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, E. Seitz, R. S. McCool, F. L. Kearns, F. Acosta-Reyes, S. Maji, et al., 2021. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature chemistry* 13:963–968.
41. Lee, C. T., and R. E. Amaro, 2018. Exascale computing: A new dawn for computational biology. *Computing in Science & Engineering* 20:18–25.
42. Keal, T. W., A.-M. Elena, A. A. Sokol, K. Stoneham, M. I. Probert, C. S. Cucinotta, D. J. Willock, A. J. Logsdail, A. Zen, P. J. Hasnip, et al., 2022. Materials and Molecular Modeling at the Exascale. *Computing in Science & Engineering* 24:36–45.
43. Vennelakanti, V., A. Nazemi, R. Mehmood, A. H. Steeves, and H. J. Kulik, 2022. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Current opinion in structural biology* 72:9–17.
44. Bernardi, R. C., and P. G. Pascutti, 2012. Hybrid QM/MM molecular dynamics study of benzocaine in a membrane environment: How does a quantum mechanical treatment of both anesthetic and lipids affect their interaction. *Journal of chemical theory and computation* 8:2197–2203.
45. Melo, M. C., R. C. Bernardi, T. Rudack, M. Scheurer, C. Riplinger, J. C. Phillips, J. D. Maia, G. B. Rocha, J. V. Ribeiro, J. E. Stone, et al., 2018. NAMD goes quantum: an integrative suite for hybrid simulations. *Nature methods* 15:351–354.
46. Phillips, J. C., D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, et al., 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of chemical physics* 153:044130.
47. Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25.

48. Ribeiro, J. V., R. C. Bernardi, T. Rudack, J. E. Stone, J. C. Phillips, P. L. Freddolino, and K. Schulten, 2016. QwikMD—integrative molecular dynamics toolkit for novices and experts. *Scientific reports* 6:1–14.
49. Humphrey, W., A. Dalke, and K. Schulten, 1996. VMD: visual molecular dynamics. *Journal of molecular graphics* 14:33–38.
50. Gowers, R. J., M. Linke, J. Barnoud, T. J. Reddy, M. N. Melo, S. L. Seyler, J. Domanski, D. L. Dotson, S. Buchoux, I. M. Kenney, et al., 2016. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. *In Proceedings of the 15th python in science conference. SciPy Austin, TX*, volume 98, 105.
51. Roe, D. R., and T. E. Cheatham III, 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* 9:3084–3095.
52. Nguyen, H., D. R. Roe, J. Swails, and D. A. Case, 2016. PYTRAJ: Interactive data analysis for molecular dynamics simulations. *New Brunswick, NJ: Rutgers University*.
53. McGibbon, R. T., K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, 2015. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* 109:1528 – 1532.
54. Grant, B. J., L. Skjærven, and X.-Q. Yao, 2021. The Bio3D packages for structural bioinformatics. *Protein Science* 30:20–30.
55. Cui, Q., and I. Bahar, 2005. Normal mode analysis: theory and applications to biological and chemical systems. CRC press.
56. Louet, M., M. Casiraghi, M. Damian, M. G. Costa, P. Renault, A. A. Gomes, P. R. Batista, C. M’Kadmi, S. Mary, S. Cantel, et al., 2021. Concerted conformational dynamics and water movements in the ghrelin G protein-coupled receptor. *Elife* 10:e63201.
57. Melo, M. C. R., R. C. Bernardi, C. de la Fuente-Nunez, and Z. Luthey-Schulten, 2020. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *The Journal of Chemical Physics* 153:134104.
58. Gheeraert, A., L. Pacini, V. S. Batista, L. Vuillon, C. Lesieur, and I. Rivalta, 2019. Exploring allosteric pathways of a v-type enzyme with dynamical perturbation networks. *The Journal of Physical Chemistry B* 123:3452–3461.
59. Sener, M., S. Levy, J. E. Stone, A. Christensen, B. Isralewitz, R. Patterson, K. Borkiewicz, J. Carpenter, C. N. Hunter, Z. Luthey-Schulten, et al., 2021. Multiscale modeling and cinematic visualization of photosynthetic energy conversion processes from electronic to cell scales. *Parallel computing* 102:102698.
60. Stone, J. E., K. L. Vandivort, and K. Schulten, 2013. GPU-accelerated molecular visualization on petascale supercomputing platforms. *In Proceedings of the 8th International Workshop on Ultrascale Visualization*. 1–8.
61. Fernández, I. S., X.-C. Bai, T. Hussain, A. C. Kelley, J. R. Lorsch, V. Ramakrishnan, and S. H. Scheres, 2013. Molecular architecture of a eukaryotic translational initiation complex. *Science* 342:1240585.
62. Bridges, H. R., J. G. Fedor, J. N. Blaza, A. Di Luca, A. Jusupow, O. D. Jarman, J. J. Wright, A.-N. A. Agip, A. P. Gamiz-Hernandez, M. M. Roessler, et al., 2020. Structure of inhibitor-bound mammalian complex I. *Nature communications* 11:1–11.
63. Gupta, C., U. Khaniya, C. K. Chan, F. Dehez, M. Shekhar, M. Gunner, L. Sazanov, C. Chipot, and A. Singharoy, 2020. Charge transfer and chemo-mechanical coupling in respiratory complex I. *Journal of the American Chemical Society* 142:9220–9230.
64. Akey, C. W., D. Singh, C. Ouch, I. Echeverria, I. Nudelman, J. M. Varberg, Z. Yu, F. Fang, Y. Shi, J. Wang, et al., 2022. Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* 185:361–378.
65. Roh, S.-H., M. Shekhar, G. Pintilie, C. Chipot, S. Wilkens, A. Singharoy, and W. Chiu, 2020. Cryo-EM and MD infer water-mediated proton transport and autoinhibition mechanisms of Vo complex. *Science Advances* 6:eabb9605.
66. Goh, B. C., J. A. Hadden, R. C. Bernardi, A. Singharoy, R. McGreevy, T. Rudack, C. K. Cassidy, and K. Schulten, 2016. Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annual Review of Biophysics* 45:253.
67. Cassidy, C. K., B. A. Himes, Z. Luthey-Schulten, and P. Zhang, 2018. CryoEM-based hybrid modeling approaches for structure determination. *Current opinion in microbiology* 43:14–23.
68. Trabuco, L. G., E. Villa, E. Schreiner, C. B. Harrison, and K. Schulten, 2009. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* 49:174–180.
69. Kovacs, J. A., V. E. Galkin, and W. Wriggers, 2018. Accurate flexible refinement of atomic models against

- medium-resolution cryo-EM maps using damped dynamics. *BMC structural biology* 18:1–11.
70. Nierzwicki, Ł., and G. Palermo, 2021. Molecular dynamics to predict cryo-EM: Capturing transitions and short-lived conformational states of biomolecules. *Frontiers in Molecular Biosciences* 8:641208.
 71. Shekhar, M., G. Terashi, C. Gupta, D. Sarkar, G. Debussche, N. J. Sisco, J. Nguyen, A. Mondal, J. Vant, P. Fromme, et al., 2021. CryoFold: Determining protein structures and data-guided ensembles from cryo-EM density maps. *Matter* 4:3195–3216.
 72. Bonomi, M., C. Camilloni, A. Cavalli, and M. Vendruscolo, 2016. Metainference: A Bayesian inference method for heterogeneous systems. *Science advances* 2:e1501177.
 73. Ridgway, D., G. Broderick, A. Lopez-Campistrous, M. Ru’aini, P. Winter, M. Hamilton, P. Boulanger, A. Kovalenko, and M. J. Ellison, 2008. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophysical journal* 94:3748–3759.
 74. McGuffee, S. R., and A. H. Elcock, 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS computational biology* 6:e1000694.
 75. Oliveira Bortot, L., Z. Bashardanesh, and D. Van der Spoel, 2019. Making soup: preparing and validating models of the bacterial cytoplasm for molecular simulation. *Journal of Chemical Information and Modeling* 60:322–331.
 76. Singharoy, A., C. Maffeo, K. H. Delgado-Magnero, D. J. Swainsbury, M. Sener, U. Kleinekathöfer, J. W. Vant, J. Nguyen, A. Hitchcock, B. Isralewitz, et al., 2019. Atoms to phenotypes: molecular design principles of cellular energy metabolism. *Cell* 179:1098–1111.
 77. Gillespie, D. T., A. Hellander, and L. R. Petzold, 2013. Perspective: Stochastic algorithms for chemical kinetics. *The Journal of chemical physics* 138:05B201_1.
 78. Holehouse, J., Z. Cao, and R. Grima, 2020. Stochastic modeling of autoregulatory genetic feedback loops: A review and comparative study. *Biophysical Journal* 118:1517–1525.
 79. Thornburg, Z. R., D. M. Bianchi, T. A. Brier, B. R. Gilbert, T. M. Earnest, M. C. Melo, N. Safronova, J. P. Sáenz, A. T. Cook, K. S. Wise, et al., 2022. Fundamental behaviors emerge from simulations of a living minimal cell. *Cell* 185:345–360.
 80. Drawert, B., A. Hellander, B. Bales, D. Banerjee, G. Belle-sia, B. J. Daigle Jr, G. Douglas, M. Gu, A. Gupta, S. Hellander, et al., 2016. Stochastic simulation service: bridging the gap between the computational expert and the biologist. *PLoS computational biology* 12:e1005220.
 81. Roberts, E., J. E. Stone, and Z. Luthey-Schulten, 2013. Lattice microbes: High-performance stochastic simulation method for the reaction-diffusion master equation. *Journal of computational chemistry* 34:245–255.
 82. Stevens, A. R., S. Bellstedt, P. J. Elahi, and M. T. Murphy, 2020. The imperative to reduce carbon emissions in astronomy. *Nature Astronomy* 4:843–851.
 83. Gibney, E., et al., 2022. How to shrink AI’s ballooning carbon footprint. *Nature* 607:648–648.
 84. Outeiral, C., M. Strahm, J. Shi, G. M. Morris, S. C. Benjamin, and C. M. Deane, 2021. The prospects of quantum computing in computational molecular biology. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 11:e1481.