



Editorial

Special Issue: “Advanced Research on Molecular Modeling of Protein Structure and Functions”

Maria G. Khrenova ^{1,2}

¹ Chemistry Department, Lomonosov Moscow State University, 119991 Moscow, Russia; khrenovamg@my.msu.ru

² Emanuel Institute of Biochemical Physics of Russian Academy of Sciences, 119334 Moscow, Russia

Recent developments in computer technologies, software and methods have made molecular modeling a powerful tool in experimental studies of biomolecular systems, and in their rational modification. GPU-accelerated molecular dynamics can operate with simulation times of microseconds, allowing researchers to track rare events. Quantum chemical calculations on both CPUs and GPUs can now deal with systems of hundreds of atoms, enabling them to be used in combined quantum mechanics/molecular mechanics biomolecular simulations. These developments have allowed researchers to deepen their understanding of the reaction mechanisms at the active sites of enzymes and photoreceptor proteins, of large conformational rearrangements of protein structures, and of the formation of protein complexes with low-molecular-weight inhibitors or macromolecular compounds. Artificial intelligence-, machine learning- and big data-based analysis methods are becoming an important part of molecular modeling. In this paper, we will present recent advances in molecular modeling related to biomacromolecules.

Significant advances in classical molecular dynamics (MD) simulations driven by improvements in hardware algorithms have revolutionized these methods and expanded the scope of their applications [1]. In the 1990s, small peptides were simulated for several nanoseconds, which was considered as a result significant enough for publication [2]. Later, development of CPU-based supercomputers expanded the range of tasks that can be solved by protein and even protein complex simulations. Still, the trajectory lengths were limited to tens of nanoseconds [3]. Development of specialized supercomputers considerably contributed to developments in this field [4,5]; however, these resources were limited. Emergence of high-performance yet affordable GPUs and further development of software optimized for novel architectures led to a revolution in classical MD simulations. Now, computations of MD trajectories of hundreds of nanoseconds or microseconds for model systems composed of hundreds of thousands of atoms are routine tasks. This allows researchers to track large conformational changes and existence of different states of entire proteins.

Development of enhanced sampling methods has further extended the possible applications of MD methods. Many of them require explicit definitions of so-called collective variables, which are usually functions of structural parameters, for example, their linear combinations [6,7]. The most frequently used methods are metadynamics and umbrella sampling. Umbrella sampling presumes addition of an external potential, usually harmonic, centered at a certain value of the collective variable, and analyzes the equilibrium distribution of states [8–10]. In contrast, metadynamics utilizes nonequilibrium sampling, allowing researchers to reconstruct an equilibrium free-energy landscape [11]. Both of these methods are frequently utilized to describe transitions between different states, particularly those actively applied to study binding processes [12–15].



Received: 23 July 2025

Accepted: 13 August 2025

Published: 16 August 2025

Citation: Khrenova, M.G. Special Issue: “Advanced Research on Molecular Modeling of Protein Structure and Functions”. *Int. J. Mol. Sci.* **2025**, *26*, 7916. <https://doi.org/10.3390/ijms26167916>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Large-scale MD simulations produce gigabytes of atomic coordinates and thousands of trajectory frames that cannot be analyzed by visual inspection and evaluation of simple geometry parameters (interatomic distances, hydrogen bonds, etc.) or using cumulative descriptors like RMSD and RMSF [16]. Valuable information can be extracted from MD trajectories using modern methods such as big data analysis and machine learning [16–23]. Kinetic analysis of transitions between different states is usually studied using Markov state modeling (MSM), which utilizes distributed MD simulation data to determine long-term behavior [24,25]. This approach requires so-called featurization and dimension reduction to a set of slow collective variables [26,27]. Recently, the variational approach for Markov processes (VAMP) was proposed to develop a deep learning framework for molecular kinetics using neural networks, called VAMPnet [28,29]. It combines the whole data processing pipeline into a single end-to-end framework and encodes the entire mapping from molecular coordinates to Markov states, providing easily interpretable kinetic models with few states.

Data clustering is frequently applied to large MD trajectories and obtains ensembles with different conformations [16]; MDTraj [30] and EnGens [31] allow researchers to perform this analysis. Interestingly, analyzing sets of static geometry parameters and their changes along trajectories can enable researchers to distinguish fragments of large biomolecular systems that are characterized by correlated motions, making this a powerful method for characterizing interactions in enzyme–inhibitor complexes [32–34].

QM/MM methods are mostly used to study chemical and photochemical/photophysical processes in biomacromolecular systems [35–37]. Initially, QM/MM simulations were utilized to locate stationary points on potential energy surfaces (PESs) and reconstruct reaction mechanisms. Today, QM/MM is also utilized for molecular dynamics simulations [38–45]. This considerably improves our understanding and provides a more realistic description of reaction mechanisms as proteins and their active sites have numerous conformations. Therefore, dealing with a particular PES minimum can lead to an incorrect qualitative understanding of the reaction mechanism. For example, a high-energy minimum can be practically unpopulated but serves as an initial state of energetically favorable reaction pathway. Different studies have demonstrated the existence of reactive and non-reactive reagent states that can only be tracked in molecular dynamics simulations [46,47].

Recent years have seen the emergence of an interdisciplinary field combining artificial intelligence (AI), machine learning (ML) and molecular modeling. Many of these methods are based on sequence inputs and utilize complex algorithms trained on large databases for predictions of 3D structures [48,49] and required properties [50].

Computer and software developments in the last decade have considerably increased the power of molecular modeling and its application in solving a wide variety of tasks. Big data analysis further enhances the potential of molecular modeling due to its ability to extract structural and dynamic data that cannot be obtained through visual analysis. Utilization of AI and ML algorithms in molecular modeling represents the next step in further developing this field of research.

Funding: This research was funded by the Russian Science Foundation (grant number 19-73-20032).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Brooks, C.L.; MacKerell, A.D.; Post, C.B.; Nilsson, L. Biomolecular dynamics in the 21st century. *Biochim. Biophys. Acta-Gen. Subj.* **2024**, *1868*, 130534. [[CrossRef](#)] [[PubMed](#)]
2. Karpen, M.E.; Tobias, D.J.; Brooks, C.L. Statistical clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of YPGDV. *Biochem.* **1993**, *32*, 412–420. [[CrossRef](#)]

3. Khrenova, M.; Domratcheva, T.; Grigorenko, B.; Nemukhin, A. Coupling between the BLUF and EAL domains in the blue light-regulated phosphodiesterase BlrP1. *J. Mol. Model.* **2011**, *17*, 1579–1586. [[CrossRef](#)]
4. Kuskin, J.S.; Young, C.; Grossman, J.P.; Batson, B.; Deneroff, M.M.; Dror, R.O.; Shaw, D.E. Incorporating flexibility in Anton, a specialized machine for molecular dynamics simulation. In Proceedings of the 2008 IEEE 14th International Symposium on High Performance Computer Architecture, Salt Lake City, UT, USA, 16–20 February 2008; pp. 343–354.
5. Shim, K.S.; Greskamp, B.; Towles, B.; Edwards, B.; Grossman, J.P.; Shaw, D.E. The Specialized High-Performance Network on Anton 3. In Proceedings of the 2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Seoul, Republic of Korea, 2–6 April 2022; pp. 1211–1223.
6. Bhakat, S. Collective variable discovery in the age of machine learning: Reality, hype and everything in between. *RSC Adv.* **2022**, *12*, 25010–25024. [[CrossRef](#)]
7. Fiorin, G.; Klein, M.L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362. [[CrossRef](#)]
8. Aho, N.; Groenhof, G.; Buslaev, P. Do All Paths Lead to Rome? How Reliable is Umbrella Sampling Along a Single Path? *J. Chem. Theory Comput.* **2024**, *20*, 6674–6686. [[CrossRef](#)]
9. You, W.; Tang, Z.; Chang, C.A. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J. Chem. Theory Comput.* **2019**, *15*, 2433–2443. [[CrossRef](#)]
10. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932–942. [[CrossRef](#)]
11. Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212. [[CrossRef](#)]
12. Zhang, C.; Zhang, R.; Qi, Y.; Wen, X.; Sun, J.; Xiao, P. Exploring the Binding Mechanism of ADGRG2 Through Metadynamics and Biochemical Analysis. *Int. J. Mol. Sci.* **2024**, *26*, 167. [[CrossRef](#)] [[PubMed](#)]
13. González-Periañez, S.; Hernández-Rosas, F.; López-Rosas, C.A.; Ramos-Morales, F.R.; Zurutuza-Lorméndez, J.I.; García-Rodríguez, R.V.; Olivares-Romero, J.L.; Ramos-Hernández, R.R.; Bravo-Espinoza, I.; Vidal-Limon, A.; et al. Multistage Molecular Simulations, Design, Synthesis, and Anticonvulsant Evaluation of 2-(Isoindolin-2-yl) Esters of Aromatic Amino Acids Targeting GABAA Receptors via π - π Stacking. *Int. J. Mol. Sci.* **2025**, *26*, 6780. [[CrossRef](#)]
14. Coffman, R.E.; Bidone, T.C. Application of Funnel Metadynamics to the Platelet Integrin α IIb β 3 in Complex with an RGD Peptide. *Int. J. Mol. Sci.* **2024**, *25*, 6580. [[CrossRef](#)] [[PubMed](#)]
15. Fedotova, M.V.; Chuev, G.N. The Three-Dimensional Reference Interaction Site Model Approach as a Promising Tool for Studying Hydrated Viruses and Their Complexes with Ligands. *Int. J. Mol. Sci.* **2024**, *25*, 3697. [[CrossRef](#)] [[PubMed](#)]
16. Kulakova, A.M.; Khrenova, M.G.; Zvereva, M.I.; Polyakov, I.V. Domain Mobility in the ORF2p Complex Revealed by Molecular Dynamics Simulations and Big Data Analysis. *Int. J. Mol. Sci.* **2024**, *26*, 73. [[CrossRef](#)] [[PubMed](#)]
17. Brownless, A.-L.R.; Rheume, E.; Kuo, K.M.; Kamerlin, S.C.L.; Gumbart, J.C. Using Machine Learning to Analyze Molecular Dynamics Simulations of Biomolecules. *J. Phys. Chem. B* **2025**, *129*, 5375–5385. [[CrossRef](#)]
18. Fleetwood, O.; Kasimova, M.A.; Westerlund, A.M.; Delemotte, L. Molecular Insights from Conformational Ensembles via Machine Learning. *Biophys. J.* **2020**, *118*, 765–780. [[CrossRef](#)]
19. Wang, Y.; Lamim Ribeiro, J.M.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139–145. [[CrossRef](#)]
20. Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390. [[CrossRef](#)]
21. Marchetti, F.; Moroni, E.; Pandini, A.; Colombo, G. Machine Learning Prediction of Allosteric Drug Activity from Molecular Dynamics. *J. Phys. Chem. Lett.* **2021**, *12*, 3724–3732. [[CrossRef](#)]
22. Jackson, N.E.; Savoie, B.M.; Statt, A.; Webb, M.A. Introduction to Machine Learning for Molecular Simulation. *J. Chem. Theory Comput.* **2023**, *19*, 4335–4337. [[CrossRef](#)]
23. Prašnikar, E.; Ljubič, M.; Perdih, A.; Borišek, J. Machine learning heralding a new development phase in molecular dynamics simulations. *Artif. Intell. Rev.* **2024**, *57*, 102. [[CrossRef](#)]
24. Scherer, M.K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542. [[CrossRef](#)]
25. Harrigan, M.P.; Sultan, M.M.; Hernández, C.X.; Husic, B.E.; Eastman, P.; Schwantes, C.R.; Beauchamp, K.A.; McGibbon, R.T.; Pande, V.S. MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **2017**, *112*, 10–15. [[CrossRef](#)]
26. Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2113533118. [[CrossRef](#)]
27. Chen, W.; Sidky, H.; Ferguson, A.L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 214114. [[CrossRef](#)]
28. Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5. [[CrossRef](#)] [[PubMed](#)]

29. Mardt, A.; Hempel, T.; Clementi, C.; Noé, F. Deep learning to decompose macromolecules into independent Markovian domains. *Nat. Commun.* **2022**, *13*, 7101. [\[CrossRef\]](#) [\[PubMed\]](#)
30. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.-P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Conev, A.; Rigo, M.M.; Devaurs, D.; Fonseca, A.F.; Kalavadwala, H.; de Freitas, M.V.; Clementi, C.; Zanatta, G.; Antunes, D.A.; Kavraki, L.E. EnGens: A computational framework for generation and analysis of representative protein conformational ensembles. *Brief. Bioinform.* **2023**, *24*, bbad242. [\[CrossRef\]](#)
32. Cowan, B.S.; Thayer, K.M. Network Theory Analysis of Allosteric Drug-Rescue Mechanisms in the Tumor Suppressor Protein p53 Y220C Mutant. *Int. J. Mol. Sci.* **2025**, *26*, 6884. [\[CrossRef\]](#)
33. Richter, M.; Khrenova, M.; Kazakova, E.; Riabova, O.; Egorova, A.; Makarov, V.; Schmidtke, M. Dynamic features of virus protein 1 and substitutions in the 3-phenyl ring determine the potency and broad-spectrum activity of capsid-binding pyrazolo[3,4-d]pyrimidines against rhinoviruses. *Antiviral Res.* **2024**, *231*, 105993. [\[CrossRef\]](#)
34. Bernetti, M.; Bosio, S.; Bresciani, V.; Falchi, F.; Masetti, M. Probing allosteric communication with combined molecular dynamics simulations and network analysis. *Curr. Opin. Struct. Biol.* **2024**, *86*, 102820. [\[CrossRef\]](#)
35. Senn, H.M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem Int. Ed.* **2009**, *48*, 1198–1229. [\[CrossRef\]](#)
36. Clemente, C.M.; Capece, L.; Martí, M.A. Best Practices on QM/MM Simulations of Biological Systems. *J. Chem. Inf. Model.* **2023**, *63*, 2609–2627. [\[CrossRef\]](#)
37. Khrenova, M.G.; Mulashkina, T.I.; Kulakova, A.M.; Polyakov, I.V.; Nemukhin, A.V. Computer Modeling of the Mechanisms of Enzymatic Reactions: Lessons from 20 Years of Practice. *Mosc. Univ. Chem. Bull.* **2024**, *79*, 86–92. [\[CrossRef\]](#)
38. Rossetti, G.; Mandelli, D. How exascale computing can shape drug design: A perspective from multiscale QM/MM molecular dynamics simulations and machine learning-aided enhanced sampling algorithms. *Curr. Opin. Struct. Biol.* **2024**, *86*, 102814. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Pederson, J.P.; McDaniel, J.G. PyDFT-QMMM: A modular, extensible software framework for DFT-based QM/MM molecular dynamics. *J. Chem. Phys.* **2024**, *161*, 034103. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Melo, M.C.R.; Bernardi, R.C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J.C.; Maia, J.D.C.; Rocha, G.B.; Ribeiro, J.V.; Stone, J.E.; et al. NAMD goes quantum: An integrative suite for hybrid simulations. *Nat. Methods* **2018**, *15*, 351–354. [\[CrossRef\]](#)
41. Bösel, L.; Thürlmann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658. [\[CrossRef\]](#)
42. Li, C.; Chan, G.K.-L. Accurate QM/MM Molecular Dynamics for Periodic Systems in GPU4PySCF with Applications to Enzyme Catalysis. *J. Chem. Theory Comput.* **2025**, *21*, 803–816. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q. QM/MM free energy simulations: Recent progress and challenges. *Mol. Simul.* **2016**, *42*, 1056–1078. [\[CrossRef\]](#)
44. Yagi, K.; Gunst, K.; Shiozaki, T.; Sugita, Y. High-performance QM/MM Enhanced Sampling Molecular Dynamics Simulations with GENESIS SPDYN and QSimulate-QM. *J. Chem. Theory Comput.* **2025**, *21*, 4016–4029. [\[CrossRef\]](#)
45. Zhou, Y.; Wang, S.; Li, Y.; Zhang, Y. Born–Oppenheimer Ab Initio QM/MM Molecular Dynamics Simulations of Enzyme Reactions. *Methods Enzymol.* **2016**, *577*, 105–118. [\[CrossRef\]](#)
46. Neves, R.P.P.; Fernandes, P.A.; Ramos, M.J. Role of Enzyme and Active Site Conformational Dynamics in the Catalysis by α -Amylase Explored with QM/MM Molecular Dynamics. *J. Chem. Inf. Model.* **2022**, *62*, 3638–3650. [\[CrossRef\]](#)
47. Polyakov, I.V.; Meteleshko, Y.I.; Mulashkina, T.I.; Varentsov, M.I.; Krinitskiy, M.A.; Khrenova, M.G. Substrate Activation Efficiency in Active Sites of Hydrolases Determined by QM/MM Molecular Dynamics and Neural Networks. *Int. J. Mol. Sci.* **2025**, *26*, 5097. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Bertoline, L.M.F.; Lima, A.N.; Krieger, J.E.; Teixeira, S.K. Before and after AlphaFold2: An overview of protein structure prediction. *Front. Bioinforma.* **2023**, *3*, 1120370. [\[CrossRef\]](#) [\[PubMed\]](#)
49. AlphaFold3 — why did Nature publish it without its code? *Nature* **2024**, *629*, 728. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Serebrennikova, M.; Grafskaa, E.; Maltsev, D.; Ivanova, K.; Bashkurov, P.; Kornilov, F.; Volynsky, P.; Efremov, R.; Bocharov, E.; Lazarev, V. TriPEP-CPP: Algorithm for Predicting the Properties of Peptide Sequences. *Int. J. Mol. Sci.* **2024**, *25*, 6869. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.