THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB
Article

CHARMM at 45: Enhancements in Accessibility, Functionality, and Speed

Published as part of The Journal of Physical Chemistry B special issue "Recent Advances in Simulation Software and Force Fields".

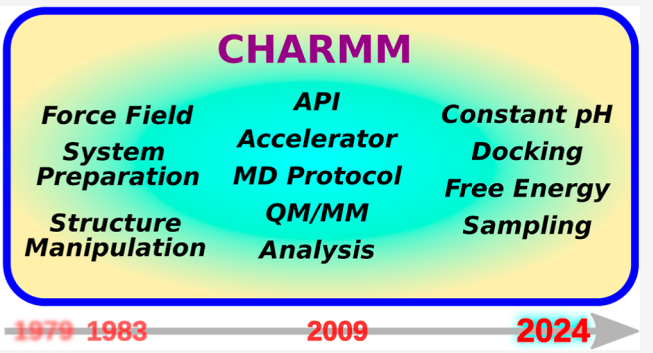
Wonmuk Hwang,* Steven L. Austin, Arnaud Blondel, Eric D. Boittier, Stefan Boresch, Matthias Buck, Joshua Buckner, Amedeo Caflisch, Hao-Ting Chang, Xi Cheng, Yeol Kyo Choi, Jhih-Wei Chu, Michael F. Crowley, Qiang Cui, Ana Damjanovic, Yuqing Deng, Mike Devereux, Xinqiang Ding, Michael F. Feig, Jiali Gao, David R. Glowacki, James E. Gonzales, II, Mehdi Bagerhi Hamaneh, Edward D. Harder, Ryan L. Hayes, Jing Huang, Yandong Huang, Phillip S. Hudson, Wonpil Im, Shahidul M. Islam, Wei Jiang, Michael R. Jones, Silvan Käser, Fiona L. Kearns, Nathan R. Kern, Jeffery B. Klauda, Themis Lazaridis, Jinhyuk Lee, Justin A. Lemkul, Xiaorong Liu, Yun Luo, Alexander D. MacKerell, Jr., Dan T. Major, Markus Meuwly, Kwangho Nam, Lennart Nilsson, Victor Ovchinnikov, Emanuele Paci, Soohyung Park, Richard W. Pastor, Amanda R. Pittman, Carol Beth Post, Samarjeet Prasad, Jingzhi Pu, Yifei Qi, Thenmalarchelvi Rathinavelan, Daniel R. Roe, Benoit Roux, Christopher N. Rowley, Jana Shen, Andrew C. Simmonett, Alexander J. Sodt, Kai Töpfer, Meenu Upadhyay, Arjan van der Vaart, Luis Itza Vazquez-Salazar, Richard M. Venable, Luke C. Warrensford, H. Lee Woodcock, Yujin Wu, Charles L. Brooks, III,* Bernard R. Brooks,* and Martin Karplus

Cite This: https://doi.org/10.1021/acs.jpcb.4c04100

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Since its inception nearly a half century ago, CHARMM has been playing a central role in computational biochemistry and biophysics. Commensurate with the developments in experimental research and advances in computer hardware, the range of methods and applicability of CHARMM have also grown. This review summarizes major developments that occurred after 2009 when the last review of CHARMM was published. They include the following: new faster simulation engines, accessible user interfaces for convenient workflows, and a vast array of simulation and analysis methods that encompass quantum mechanical, atomistic, and coarse-grained levels, as well as extensive coverage of force fields. In addition to providing the current snapshot of the CHARMM development, this review may serve as a starting point for exploring relevant theories and computational methods for tackling contemporary and emerging problems in biomolecular systems. CHARMM is freely available for academic and nonprofit research at https://academiccharmm.org/program.



CONTENTS

Received: June 20, 2024
Revised: August 15, 2024
Accepted: August 22, 2024
Published: September 20, 2024

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

1. INTRODUCTION

CHARMM, the program for simulation and modeling of biomolecular systems, is now more than 40 years in continuous development and use. It is freely available to academic and government laboratory users, and is fast, as we describe below due to recent advances in GPU acceleration. From its earliest incarnations as Gandalf, renamed to HARMM (Harvard Macromolecular Mechanics), and finally CHARMM (Chemistry at Harvard Macromolecular Mechanics), it has provided a working framework for the exploration of biomolecular structure−function−dynamics relationships, beginning with the first molecular dynamics (MD) simulation of the small protein pancreatic trypsin inhibitor.[1] For an interesting perspective on the evolution of simulation methods and the applications of statistical mechanics to the study of biological molecules, we point readers to a recent review.[2]

CHARMM provides many practical and functional features that distinguish it from other programs that have evolved for similar purposes. Probably the most significant is that CHARMM was created as, and has remained, a repository for many of the trend-setting methods and models. It now comprises over 1,170,000 lines of code (modular Fortran 90, C, C++, CUDA, OpenCL, and Python) encapsulating its extensive functionality. An equally important and distinctive feature that has been integral to the software since its earliest days, is an interpreted language as its command parser, enabling "programs" to be written in "CHARMM-language" in contrast to other programs in this area whereby one prepares an input "script" describing the particular (one-pass) calculation one wishes to run. This feature greatly facilitates testing and prototyping of many of the statistical mechanical methods and techniques that have been integrated into CHARMM. While much of this has been described in earlier works and will not be further elaborated here, it is important to recognize this fundamental differentiator of CHARMM and other programs utilized in the field.

In this review, we provide an update to the developments that have occurred in CHARMM since it was last described in 2009.[3] We refer readers to the previous two papers describing CHARMM for basic organization of the program and other functionality incorporated prior to 2009.[3,4] Among the key new developments during the past 15 years, the most significant are the establishment of GPU-accelerated kernels to perform many of the unique calculations available in CHARMM. We describe the three published CHARMM accelerator engines in Sections 2.1 (CHARMM/OpenMM API), 2.2 (CHARMM/DOMDEC), 2.3 (CHARMM/BLaDE), and 2.4 (apoCHARMM). Each of the former three platforms are fully integrated with CHARMM and support a significant range of CHARMM functionality, and thereby provide powerful platforms for establishing complex simulation workflows utilizing CHARMM scripting language. apoCHARMM is a new GPU accelerator that is currently being developed. Except for DOMDEC, these engines provide performance comparable to any existing GPU-based biomolecular simulation code. Thus, CHARMM performance is on par with other codes while the accessibility of methods is typically richer.

Beyond the advanced simulation engines, significant efforts to integrate CHARMM into modern workflows have advanced, as described in Sections 3.1 (pyCHARMM), 3.2 (crimm), and 3.3 (CHARMM-GUI). CHARMM-GUI continues to mature and is a vital service to the community through its system and simulation setup facilities, providing input scripts for simulations for a number of current biomolecular MD simulation packages, including CHARMM, OpenMM, Amber, and GROMACS (Section 3.3). Section 3.1 on pyCHARMM describes recent efforts to use CHARMM functionality within the context of Python language. This enables the straightforward integration of CHARMM and the immense base of Python modules. Finally, crimm is a Python-based package that integrates many simulation preparation tasks, e.g., building of missing residues and loops, choosing protonation states appropriate for a given pH, patching to represent disulfide and other system modifications and solvation. These tools and methods of CHARMM provide an essential platform for modern simulation and modeling workflows.

New docking methods and procedures are described in Section 4, followed by an update of the free energy (Section 5) and constant pH (Section 6) methods. We then discuss the new enhanced sampling and transition path methods (Section 7). CHARMM has supported a range of reactive, implicit and coarse-grained models for simulation of biorelated systems throughout its history and new developments are discussed in Section 8. System-specific and specialized restraint methods that have been developed and implemented in CHARMM recently are in Section 9.

CHARMM's force fields (FFs) are corner posts of molecular simulations throughout the field. Section 10 gives updates to the CHARMM fixed-charge and polarizable FFs. Following is a discussion of the recent approaches and methods for quantum mechanics and molecular mechanics (QM/MM) simulations that are integrated into CHARMM (see Section 11). Lastly, we describe new methods, procedures and analysis tools that have been integrated into CHARMM in Section 12.

Through this review, we hope to convey the immense base of methods and models that are available and supported in CHARMM. Additionally, we hope that readers will use this review as an entry to the growing online repositories for tutorials, advanced simulation methods, and templates.

2. CHARMM ACCELERATOR ENGINES

Since the previous review of CHARMM functionality,[3] there has been a significant growth in utilizing new generation of processors, especially GPUs, to accelerate MD and related modeling tasks within biomolecular simulation methods. The CHARMM development community has embraced this effort by developing specialized GPU accelerator and highly parallel kernels, or adaptor APIs that support a range of the extensive functionality available in CHARMM, from free energy methods and constant pH simulation techniques, to implicit solvent models such as generalized Born (GB) models and a host of others. These interfaces provide a straightforward means to set up systems, manipulate, patch or otherwise prepare them for simulation and then simulate using GPU or parallel CPU execution without creating extraneous intermediate files. This seamless interface also enables straightforward analysis and visualization of results from within the same pyCHARMM/CHARMM script, thereby providing an integrated framework for modeling, dynamics, and analysis. We

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 1. CHARMM Accelerated Engine Benchmarks on GPUs (ns/day; avg ± sd over 5 runs)[a]

	System	DHFR	APOA1	DMPG	T4L (N_s = 6)	HSP90 (N_s = 11)
GPU	Atoms	23558	92224	291168	37723	26685
Ada RTX6000	BLaDE	601.6 ± 2.4	255.6 ± 1.2	82.1 ± 0.4	-	-
	OpenMM	1052.0 ± 3.5	257.8 ± 2.8	78.0 ± 0.9	-	-
	apoCHARMM	423.5 ± 1.3	289.2 ± 1.4	95.3 ± 2.5	-	-
Ada RTX5000	BLaDE	514.3 ± 6.0	191.7 ± 0.6	63.6 ± 0.2	-	-
	OpenMM	897.3 ± 1.1	197.7 ± 2.3	62.7 ± 1.3	-	-
	apoCHARMM	323.6 ± 11.5	218.3 ± 0.7	72.3 ± 0.3	-	-
Ada RTX4500	BLaDE	436.7 ± 0.45	142.9 ± 0.2	45.6 ± 0.04	-	-
	OpenMM	730.0 ± 0.3	138.9 ± 1.2	43.6 ± 0.6	-	-
	apoCHARMM	388.6 ± 0.7	152.2 ± 0.7	51.3 ± 0.2	-	-
RTX A6000	BLaDE	355.9 ± 2.7	124.5 ± 0.2	40.3 ± 0.4	-	-
	OpenMM	656.0 ± 6.5	129.8 ± 1.6	39.5 ± 0.5	-	-
	apoCHARMM	295.8 ± 0.7	133.8 ± 1.1	42.6 ± 0.3	-	-
RTX A5500	BLaDE	333.5 ± 1.0	111.8 ± 4.7	37.9 ± 0.3	-	-
	OpenMM	594.3 ± 2.7	113.4 ± 0.6	31.2 ± 0.1	-	-
	apoCHARMM	277.2 ± 5.6	135.3 ± 2.4	38.5 ± 0.2	-	-
A100	BLaDE	402.9 ± 4.7	149.4 ± 0.9	54.0 ± 0.2	208.8 ± 0.3	246.8 ± 0.4
	DOMDEC	106.3 ± 0.7	-	9.07 ± 0.06	33.5 ± 0.0	46.4 ± 0.3
	OpenMM	558.5 ± 8.6	121.7 ± 2.3	36.2 ± 0.6	-	-
	apoCHARMM	276.1 ± 1.5	172.6 ± 1.3	63.6 ± 0.3	-	-

[a]DHFR, APOA1, and DMPG are for NVT simulations with 9-Å real-space non-bonded cutoff distance. T4L and HSP90 are for λ-dynamics in NPT ensemble and with 10-Å real-space non-bonded cutoff distance. N_s in parentheses is the number of λ variables used (*cf.*, Eq. 4). PME was used to account for long-range electrostatic interactions.[24] In all simulations, the integration time step was 2 fs.

present below in chronological order, the accelerated performance platforms that are integrated into CHARMM. As indicated in Table 1, the performance of CHARMM through the GPU-accelerated APIs compares well with those observed in other GPU-accelerated packages such as pmemd.cuda,[5] GROMACS,[6] and NAMD.[7] CHARMM/pyCHARMM (Section 3.1) also offers a range of developed and developing interfaces to meet needs across a breadth of the methodological application areas.

2.1. The OpenMM API. The first GPU-accelerated engine coupled to CHARMM comprised a FORTRAN-90 API that takes advantage of the significant developments of OpenMM.[8] This interface provides direct calls to OpenMM functionality for MD, energy minimization, and free energy methods. A host of restraints existing in CHARMM are also implemented using OpenMM's custom forces.[8] This effort, spearheaded by Michael Garrahan and Charles Brooks, first appeared in CHARMM release c37b1. A key advantage of using OpenMM through the CHARMM/OpenMM API is that applications that need to move between system setup and preparation, processing, and analysis can occur within a single workflow using the CHARMM interpreted command language or directly though Python with pyCHARMM.[9] Aside from the many simulation environments, restraints, heavy atom-hydrogen constraints, NVT and NPT (using either an isotropic Monte-Carlo barostat or anisotropic pressure coupling often used in membrane simulations), particle-mesh Ewald (PME), CHARMM shifting and switching methods for van der Waals and/or electrostatic interactions, this interface also supports free energy perturbation (FEP) methods utilizing fixed windowing approaches, e.g., FEP with analysis by MBAR (Section 11.11).[10−12] The CHARMM/OpenMM API provides a robust platform for integrated modeling tasks and workflows and has been utilized in the extension of the CDOCKER approaches[13−15] to parallel simulated annealing with GPU acceleration with any of the CHARMM-compatible physically based FFs, including CHARMM,[16] CHARMM General FF (CGenFF),[17,18] AMBER,[19] GAFF,[20] OPLS,[21] and the LigParGen OPLS extension for small molecules.[22] At present, not all of the CHARMM functionalities are fully implemented through the API, and a key missing element is full access to the Drude polarizable FF that has been rapidly developing over the past several years.[23] Plans are underway to provide availability through the CHARMM/OpenMM API in the near future.

2.2. DOMDEC Parallel-Scalable Platform. In 2014, a new DOMain DEComposition (DOMDEC) MD engine was introduced into CHARMM by Hynninen and Crowley.[25] It was faster both in the execution on serial and parallel CPU platforms. Serial performance was approximately two times higher than in the previous versions of CHARMM with its "fast" CPU-based options. The parallel version enabled efficient utilization up to hundreds of CPU cores.

The DOMDEC module of CHARMM served as an early platform for the development of the multisite λ-dynamics (MSλD)[26−28] and explicit-solvent constant pH MD (CpHMD)[29−31] methods. Also implemented as part of this effort were GPU-resident kernels that accelerated components of the computation and enabled partitioning of the system being studied across both CPU and GPU cores for scaling and acceleration. Finally, a GPU implementation that handles all energy calculations except for the SHAKE constraint and position propagation was implemented in the "GPU only" functionality of DOMDEC. All of CHARMM's MSλD and CpHMD functions were integrated into the "GPU-only" kernel, which was the platform on which NAMD's GPU-accelerated kernel (gpu-offload) was based as well as more recent faster engines including the newly developing apoCHARMM (Section 2.4).

2.3. BLaDE. The BLaDE (Basic LAmbda Dynamics Engine) module of CHARMM[24] was developed to optimize the speed of λ-dynamics simulations on GPUs, but it also provides a robust and accelerated platform for conventional

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

MD simulations. Previously, the DOMDEC module[25] was the fastest implementation of λ-dynamics as noted above, but the SHAKE constraint and position propagation being handled by the CPU were rate limiting. Also, DOMDEC performed suboptimally on smaller systems on a single GPU. BLaDE optimizes these tasks and achieves 5- to 6-fold speedup over DOMDEC. Although the CHARMM/OpenMM API (discussed above) exhibits similar performance on standard MD (Table 1), it is less suited for λ-dynamics.

Table 1 shows benchmarks for BLaDE, DOMDEC, and OpenMM through their CHARMM interfaces, and a standalone version of apoCHARMM. We have focused on previously established benchmarks as presented earlier[24] and elsewhere. DHFR is a small globular protein, ApoA1 is a solvated lipid nanodisc, DMPG is a larger lipid bilayer. T4L is a protein mutation calculation, and HSP90 is a ligand perturbation calculation. Benchmarks were repeated 5 times and run on single NVIDIA GPUs as noted in Table 1. These benchmarks demonstrate that BLaDE scales well, especially for λ-dynamics.

Since BLaDE is designed to be simple and fast, not all features present in CHARMM are available in BLaDE. For example, while much of CHARMM uses the Langevin Piston barostat,[32] BLaDE uses the Monte Carlo barostat[33,34] in constant pressure simulations, removing the overhead of computing the virial and rectifying SHAKE constraints after coordinate update. Similarly, BLaDE only includes a Langevin thermostat, and is not yet implemented in energy minimization routines, though energy calls can be made directly to BLaDE. New features continue to be added, including support for harmonic and nuclear Overhauser effect (NOE) restraints, support for non-orthogonal boxes, PME[35] or force/energy switching electrostatics,[36] and support for AMBER FF with different 1−4 scaling and improper torsion potentials.

2.4. apoCHARMM, Embracing CHARMM-Centric Functionality. apoCHARMM is a developing open source package designed specifically to support some of the distinctive methods of CHARMM absent in the CHARMM/OpenMM or CHARMM/BLaDe APIs (see Sections 2.1 and 2.3), and at the speeds provided by modern GPU architectures (Table 1). In particular, apoCHARMM in its current development supports, or plans to support:

- A complete analytic virial tensor.
- Multiple PSFs (protein structure files) simultaneously (upper limit set by the hardware resource).
- Uncommon crystal symmetries such as $P2_1$ (Section 12.4).

By accounting for the complete virial tensor, an implementation of the Langevin piston algorithm[32] for constant pressure or constant surface tension ensembles is enabled. Simultaneous support for multiple PSFs allows free energy methods modeled after the CHARMM PERT approach to be run. It also allows the enveloping distribution sampling (EDS) based method for free energies,[37−39] and for state-based CpHMD.[40,41] Support for $P2_1$ crystal symmetry[42,43] allows lipid bilayer systems to be simulated without chemical potential mismatch between upper and lower leaflets, which can be very useful when making membrane insertions[44] (Section 12.4).

A number of different integrators have been implemented for different ensembles: Velocity-Verlet and leapfrog integrators for the microcanonical ensemble, and Langevin thermostat and Nosé−Hoover integrators for the canonical ensemble. Holonomic constraints are handled using SHAKE and SETTLE algorithms. Since the virial is calculated during force calculation, the isobaric ensemble can be sampled using the Langevin piston method.[32] This is an extended ensemble method with additional degrees of freedom corresponding to pistons that are used to control the pressure. Thus, a number of ensembles are available including constant area (NPAT) and constant surface tension (NPγT).

Several methods for free energy difference calculations are implemented in apoCHARMM.[45] A unifying scheme, which depends on a variant of energy interpolation, is implemented using a composite design pattern, where forces and energies of the end states after being separately calculated are interpolated. Additionally, soft-core formulation of the van der Waals interaction[46] to calculate λ-specific energy is available. The double exponential method[47] has been implemented as well. Although it is slightly slower than the van der Waals formulation, it provides a base version of soft-core.

apoCHARMM is derived from the erstwhile GitHub package by Antti-Pekka Hynninen.[48] It is written in CUDA and modern C++ to leverage the full potential of NVIDIA GPU architectures. Additionally, it features a Pybind11-based Python interface, ensuring convenience for end-users. The codebase adheres to test driven development (TDD) principles and incorporates Catch2-based unit tests with extensive code coverage. Notably, apoCHARMM is designed as a GPU-exclusive implementation, with all aspects of MD including energy and force calculations, restraints, constraints, and integration, executed entirely on GPU. Minimizing host-GPU memory transfers, the system only necessitates such transfers during logging or trajectory saving operations. One of the modular design patterns employed is the mediator pattern that reduces dependencies between different components by mandating communication through a central mediator object. In a similar vein, loggers and integrators leverage a publisher−subscriber design pattern, facilitating the versatile reuse of different loggers with distinct integrators. Overall, apoCHARMM performance is comparable to or better than other GPU based MD engines (Table 1). Since it is optimized for larger systems, its performance is not as good for the smaller DHFR.

3. STREAMLINING CHARMM WORKFLOWS

The rich methodology and broad functionality of CHARMM, including its unique scripting language, has enabled many complex workflows to be created and tested prior to committing them to code in FORTRAN 90/C/C++.[3] These scripting capabilities have led to extensive libraries of CHARMM scripts in various forums and repositories as well as seeded the establishment of the web-based CHARMM-GUI[49] and a range of other modeling tasks, e.g, MCSS,[50] early stages of XPLOR,[51] and SILCS.[52] CHARMM scripting language, although extremely powerful, is not naturally integrable with other workflows that are convenient and widely used in the modeling of biomolecules. This realization has provided impetus for establishing a complete Python interface to CHARMM's full range of functionality and efforts to facilitate the utilization of CHARMM and pyCHARMM, namely crimm and CHARMM-GUI.

3.1. pyCHARMM. Efforts were initiated in the group of Charles Brooks to develop CHARMM callable functionality though a Python interface and APIs, called pyCHARMM.[9] It

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

enables CHARMM variables and data structures to be explored and used, and in some instances manipulated at the Python level, providing the means of creating complex workflows that integrate and extend tools built in Python for numerical and graphical tasks. Native Python functions and modules complement and extend the already rich landscape of CHARMM functionalities. Examples include a framework that enables novel energy functions to be integrated with CHARMM's modeling tools through Python callable routines available in Python, CUDA, and OpenCL, as well as utilize machine learned functions such as TORCH-ANI[53] and PhysNet[54] for energy and force calculation. Analogous 'hooks' are built into the CHARMM dynamics engine. Graphical engines are also readily integrated into pyCHARMM for rapid visualization of simulation models and results. Loosely coupling tasks across many processors too is straightforward within pyCHARMM workflows using MPI frameworks such as MPI4PY and this has facilitated free energy calculations using multiscale Bennett's acceptance ratio (MBAR) and thermodynamic integration (TI) approaches, or high-throughput MSλD free energy methods (Section 5.1), string path optimization calculations (Section 7.3), replica exchange (Section 7.1), and fully automated docking workflows employing CDOCKER (Section 4.1).[13,14,55,56] pyCHARMM is integrated with the accelerated platform kernels and APIs of CHARMM/OpenMM (Section 2.1) and CHARMM/BLaDE (Section 2.3).

Since its release in early 2023,[9] two workshops have been given on integrating modeling tasks using pyCHARMM. The first focus was on general modeling tasks and methods with examples provided as Jupyter Notebooks and Python scripts (July 2022),[57] followed by an advanced workshop held in July 2023, focused on utilizing pyCHARMM for high-accuracy, high-throughput free energy calculations.[58] The Jupyter Notebooks and scripts associated with both workshops are also available through the GitHub page of Charles Brooks' lab: https://github.com/BrooksResearchGroup-UM. The release and ongoing development of pyCHARMM represent an important milestone for integrating biomolecular modeling, FFs, advanced simulation, sampling, and docking protocols, into the widely used Python programming language.

3.2. crimm. Despite the best efforts from the developers of major biomolecular modeling and simulation softwares, there exists a substantial barrier when researchers first start to learn these tools. Frustrations often arise from the unfamiliarity of command scripts and in system preparation protocols that involve multiple steps to process macromolecules, small molecule ligands, water, and ions. A number of structural preparation tools have been developed for this purpose. For example, CHARMM-GUI (Section 3.3), originally designed to prepare simulation systems for CHARMM, provides a web interface to assist users with building a system for simulation.[49,59−62] Despite convenience, it lacks scriptability and integratability. Other tools such as PDBfixer from OpenMM provide Python APIs for scripting and possible integration with other tools. However, they rely on structure files such as PDB as an intermediary to pass structural information to other software packages. The limitation of the PDB file format has rendered it insufficient in keeping complete information of a macromolecular system. Python packages such as Biotite[63] and BioPython[64] offer adequate APIs for structure manipulation and protocols for integration with other computational tools, but they are limited in utility for structural preparation for MD simulations. Functions such as building missing loop regions, building missing atoms, adding hydrogens, solvation, *etc.,* are currently absent. To address issues of scriptability and integratability in structure preparation, crimm (Chemistry with ReInvented Macromolecular Mechanics)[65] is being designed with the following software principles and aims:

1. Accurate, consistent, and complete structure information and annotations for biomolecules maintained throughout the structure preparation pipeline.
2. Intuitive object design that organizes structural entities (e.g., model, chain, residue, atom) for retrieving information and manipulating structures, thus providing greater flexibility for programming.
3. Abstraction on routines (e.g., protonation, solvation, loop building) to create high-level APIs provided in Python for a convenient and intuitive scripting on structure preparation.
4. Clear protocols and reference implementations to Adaptors (interfaces to convert between Python classes in memory) to pass structural information to other software library or platforms, where accurate and efficient transfer of data can be guaranteed.
5. Visualizations in an interactive programming environment, i.e., Jupyter Notebook, to aid examination of structures.
6. Ease of installation, free and open source, and support for all major hardware platforms to encourage adaptation.

Crimm is directly built on the BioPython library and adopts the SMCRA model (Structure, Model, Chain, Residue, and Atom) for representing structures.[64] A BioPython-based object class provides optimal classification of macromolecular chain entities (protein, RNA, DNA, oligosaccharide, *etc.*). Importantly, functions of BioPython can be directly called with crimm structural object as an argument. All structural objects can be directly visualized using NGLView[66] in a Jupyter Notebook/Lab.

Structure preparation in crimm begins by fetching structures from the RCSB[67] or AlphaFold DataBase[68] as mmCIF format files for the complete and consistent organization of information.[69] In the context of CHARMM, the topology generation functions with CHARMM naming conventions and the CHARMM C36 FF is used. Currently available routines to process initial structures from the RCSB include automated missing loop/residue and disulfide bond assignment based on the data in the mmCIF file, patching of titratable residues with protonation state assigned by using interface to PropKa,[70] and topology generation. A solvation module is under development. Adapters to pyCHARMM[9] have been implemented and crimm structures can be operated on or one can run simulations with CHARMM functions via pyCHARMM. An Adapter to RDKit[71] has also been implemented for small molecule ligands integral to PDB entries. These are created as mol objects in RDKit to guarantee they maintain the correct bond orders. Other features are currently being developed to aid structure preparation and will address interfaces and adaptors to packages such as OpenMM,[8] OpenFF,[72] and Autodock Vina.[73]

3.3. CHARMM-GUI. Since its original development in 2006,[49,59−62] CHARMM-GUI has proven to be an ideal web-based platform (https://www.charmm-gui.org) to interactively

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

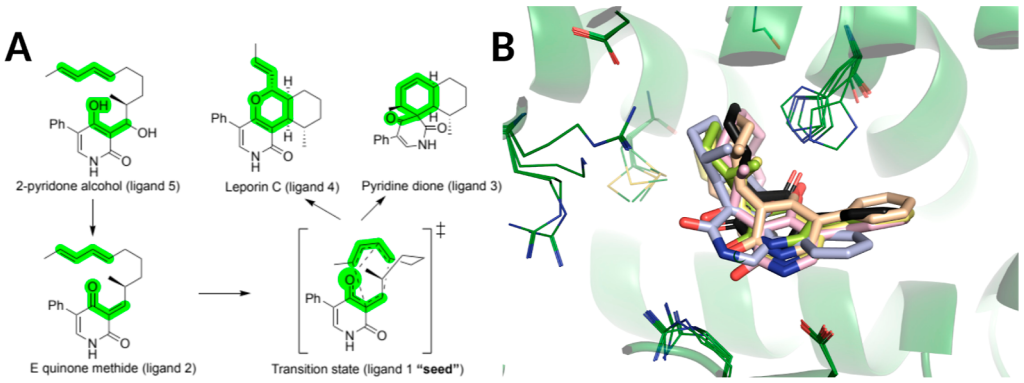


Figure 1. Main concepts of EnzyDock applied to the mechanism in the Diels–Alderase enzyme, LepI.[133,134] (A) Similar (mapped) atoms are marked in green. (B) EnzyDock docking with the transition state as a template ("seed") for docking the remaining states.

build complex systems and prepare their inputs with well-established and reproducible simulation protocols for widely used simulation packages such as CHARMM, AMBER,[74] Desmond,[75] GENESIS,[76] GROMACS,[77] LAMMPS,[78] NAMD,[79] OpenMM,[8] and Tinker.[80] CHARMM-GUI has been widely adopted for various purposes and it now contains more than 20 modules designed to set up a broad range of molecular simulation systems.[81−83] CHARMM-GUI also provides educational resources including online lecture materials, an online user forum, and workshops. Its archives support scientific reproducibility by providing the lipid conformation library[59,61,62] used in membrane generation, prebuilt COVID-19 systems,[81,84,85] prebuilt membrane complexes,[82,83] and a searchable CHARMM small molecule library (CSML). Many original modules were developed as an in-house effort, but close collaborations with the developers of CHARMM and other simulation packages have been established for adding newer modules.[86−88]

The philosophy in CHARMM-GUI development is less about providing the nuts and bolts of molecular modeling, but instead it focuses on helping users to achieve a task, including building membrane systems,[61,62,89−94] modifying and solvating proteins,[95,96] characterizing protein−ligand interactions,[97−104] or modeling complex carbohydrates[105−107] via a streamlined interface.[108−116] This makes CHARMM-GUI broadly accessible to users with little experience in modeling tools while remaining useful to experts, especially for batch generation of systems.

CHARMM-GUI development is not only guided by requests from general users and experts, but also in response to an emerging need for a unified platform to prepare and execute various advanced simulation approaches developed in diverse simulation communities and packages.[113,117−120] In addition to building complex molecular systems, CHARMM-GUI also assists with preparing input files for both general and advanced modeling and simulation tasks.

4. DOCKING METHODS

4.1. CDOCKER. First introduced in 2003,[55] CDOCKER provides an integrated CHARMM-based scripting framework for small molecule-receptor docking studies. It employs a numerical grid-based representation for the van der Waals and electrostatic interactions utilizing a fully molecular mechanics (MM)-based FF representation of the interactions.[55] CDOCKER utilizes conformational search based on simulated annealing, and it is also compatible with enhanced sampling search approaches such as self-guided Langevin dynamics.[58] It has been used in a broad range of applications, including early efforts in community-based docking.[121−125] In this capacity it served as a platform to explore a range of docking and scoring approaches, including some of the early flexible receptor[126] and covalent docking methods.[56]

In the past few years, CDOCKER has been significantly updated to utilize accelerated platforms such as GPUs.[13−15] While the basic philosophy has remained centered on sampling via (accelerated) simulated annealing and structured around the state-of-the-art small molecule and biomacromolecular FFs, new fast Fourier transform (FFT)-based approaches have been introduced for binding pocket and ligandable-site discovery via functional probe docking,[127] representing important hydrogen bonding by use of hydrogen-bond-specific donor−acceptor grids,[14] and hybrid sampling methods that combine simulated annealing with genetic algorithm moves.[14]

CDOCKER has also been implemented as a package within pyCHARMM (Section 3.1). In addition to providing full access to the methods available within CDOCKER, pyCHARMM greatly simplifies the workflow through use of 'best practice' parameter choices and a single callable pyCHARMM command. This enables large-scale virtual screening via a single script that integrates ligand building via RDKit and SMILES strings, parametrization of ligands with small molecule FF parameter estimators such as CGenFF,[17,18] GAFF,[20] LigParGen,[22] and OpenFF,[128] automated protein grid generation, parallel docking, clustering of results and ranking, including reranking with implicit solvent models such as GBSA/GBSW/GBMV, and FACTS.[129−132] In summary, CDOCKER is a fast, flexible and accurate GPU-accelerated molecular docking engine that can handle cases from high-throughput small probe docking to flexible receptor−ligand docking.

4.2. EnzyDock. Modeling enzyme reactions requires a carefully designed computational protocol that relies on well-established theoretical foundation. The starting point is reliable 3-dimensional (3D) structures of the substrate, product, intermediates, or transition states bound to the enzyme. EnzyDock[135] is a CHARMM-based docking program like the well-known CDOCKER[56,136,137] (Section 4.1), with emphasis on enzymes. Its main feature is mechanism-based multistate consensus docking that allows the docking of reaction substrate, intermediates, transition states, and products in a mechanistically consistent and induced-fit manner (Figure 1). EnzyDock is written as a series of CHARMM scripts (>10,000

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

lines of script code), Python codes (~3,000 lines), and shell scripts. EnzyDock is a docking-tool and it does not compute free energy profiles that can be obtained using other methods in CHARMM such as umbrella sampling (US),[138] string-based methods[139] (Section 7.3), or metadynamics.[140,141] Consensus docking in EnzyDock is achieved by applying geometric restraints implemented via NOE restraints on reaction states relative to a predetermined "seed" state, such that all states are docked with similar poses under a given user-defined threshold (Figure 1). For instance, the seed state could be a tightly bound transition state or a known inhibitor-bound state. Conversely, if unrestrained multistate docking is performed, a reaction pathfinder module identifies all matching poses along a reaction path.[142] Additional restraints such as on dihedral angles can enforce specific stereo- and regio-chemistry during docking, while positional harmonic and NOE restraints can be employed to include chemical information such as the initial cleavage site, nucleophilic attack, or ligand positions relative to key active-site residues or cofactors. Different protonation states of enzyme and cofactors during docking of different states is facilitated via CHARMM patching. Sampling of configurational space is performed using MD[3] or Monte Carlo (MC)[143] simulated annealing on a grid representing the enzyme,[55] and poses are scored using the C36[144] and CGenFF[145] FFs. Flexible residues, cofactors, and waters are treated as explicit atoms on the grid. Following ligand pose clustering, final energy minimization and scoring is performed using all-atom description of the entire system and optional refinement using a QM/MM approach[135,146] with a range of QM methods, e.g., semiempirical (SE)[147] or density functional theory (DFT)[148] (Section 11). Bulk solvation is modeled using an implicit solvation model (e.g., GB).[3]

EnzyDock has been applied to diverse systems such as terpene synthases, racemases, Diels–Alderases, phosphotriesterase,[149] and covalently bound ligands.[135,150] It has also been used along with other docking programs in a benchmark study on ligand binding in the main protease in SARS-CoV-2.[151]

From a user perspective, the enzyme must be provided as a PDB file or CHARMM PSF and CRD (coordinate) files, ligand states as PDB files or SMILES strings, and atom mapping between similar states along a reaction path must be provided by the user. Additional restraints can be provided by the user. EnzyDock is available via GitHub and has recently been implemented in CHARMM-GUI[49] (Section 3.3).

4.3. CIFDock. Accurately modeling protein and ligand flexibility is vital when using molecular docking to elucidate binding modes and predict binding affinity. Binding events may rely on "induced fit" where the ligand induces conformational changes in the protein binding site.[152] Accounting for induced fit has been shown to be critical for accurate modeling of the complex.[153,154] To this end, we developed a novel CHARMM-based induced fit docking protocol (CIFDock)[155] that employs all-atom FFs and enhanced sampling MD.

The CIFDock protocol begins with processing the protein structure through CHARMM-GUI[96] to fix bond orders, add hydrogens, and correct protonation states of the protein residues. The resulting PDB file is then fed into a series of CHARMM scripts which separate the protein, ligand, ion, and water molecules into CHARMM-compatible structure files, and they will be combined during subsequent steps. Key to the CIFDock protocol is the definition of active site residues that are mutated to alanine, to allow for a more "open" active site that can accommodate larger ligands and facilitate greater ligand conformational searching. In the final preparation step, the Confab module of OpenBabel[156] is used to generate a ligand conformational ensemble to seed initial binding pose searching.

The main docking procedure begins with the initial placement of the ligand in the active site of the protein in a random orientation. The ligands are then sampled using a 20-ps self-guided Langevin dynamics (SGLD)[157] simulation. Following this step, pairwise root-mean-square deviation (RMSD) clustering of ligand conformations is performed using the CORREL module to avoid further sampling of overlapping conformations. Each cluster is saved as a trajectory file which consists of conformations that were within a predefined cutoff radius of the cluster center. Each of the resulting protein–ligand complexes is then "backmutated" (i.e., the residues mutated to Ala are mutated back to their original residues) and a random dihedral angle-based rotamer library[158] is generated, and side chains are relaxed by a short energy minimization and SGLD simulation. Explicit water molecules and ions saved in the preparation stage are added back, and a second SGLD simulation is conducted on the active site complex.

The resulting "docked" poses are scored and ranked using a set of custom scoring functions that are based on the well-validated SWISSDOCK[159] scoring function. They are linear combinations of energy terms calculated by CHARMM, which include FF-based energies and the GMBV II implicit solvent model for solvation energy.[160] CIFDock was validated by cross-docking studies on a set of 21 pharmaceutically relevant proteins. Results obtained were comparable to, or in some cases improved upon, commercial docking programs. This can be attributed to the treatment of the ligand, active site, and explicit waters as fully flexible components during the docking procedure. Additionally, because CIFDock is based on short classical MD simulations, its computational cost is minimal.

To handle the formation of covalent bonds and allow covalent inhibitors to be screened, we integrated both MNDO and SCC-DFTB[161−165] minimizations (Section 11) into the CIFDock workflow. These minimizations together with additional dynamics simulations using positional restraints ensure adequate protein–ligand complex sampling pre- and postreaction. The covalent-based CIFDock (CovCIFDock) has been validated on a cross-dock and self-dock test set,[166,167] with an average RMSD of 1.91 and 1.89 Å, respectively, and a 76% success rate. This compares favorably with commercial covalent docking programs such as Schrödinger's CovDock-LO (Lead Optimization) that has a 74% success rate on the same test sets. The hybrid QM/MM minimizations also add little computational overhead to the docking procedure.

5. FREE ENERGY METHODS

5.1. λ-Dynamics, Multisite λ-Dynamics, and Constant pH MD. Alchemical free energy simulations are an important class of statistical mechanical methods used in computing free energy values and differences in small molecule design and refinement,[28,168,169] as well as protein design[170,171] and CpHMD simulations.[50,172] Alchemical methods determine free energy differences by simulating chemical transformations along a non-physical pathway, often using a chemical progress variable λ. λ-dynamics is a particularly efficient and scalable alchemical method that takes advantage of natural fluctuations in the systems being studied to "drive" the chemical coordinate between the desired end points, and is generalizable to

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

multidimensional chemical spaces, allowing exploration of many substituents at a site or even at multiple sites (MSλD) in a single simulation.[26,173]

For two states A and B of a molecular species (e.g., a protein or a side chain) and the environment (e.g., solvent and/or the receptor pocket), the alchemical hybrid Hamiltonian (or Lagrangian) for the λ-dynamics is

$$\begin{aligned}\mathcal{H}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda) &= (1-\lambda)U_A(\mathbf{r}_A, \mathbf{R}_e) + \lambda U_B(\mathbf{r}_B, \mathbf{R}_e) \\ &\quad + U_e(\mathbf{R}_e) + U_{\text{bias}}(\lambda) + K_A + K_B + K_e \\ &\quad + K_\lambda \\ &= \mathcal{U}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda) + U_{\text{bias}}(\lambda) + K_A + K_B \\ &\quad + K_e + K_\lambda \end{aligned} \tag{1}$$

where $U_{A/B}$ represent potential energy of A/B interacting with themselves and the environment and U_e is the potential energy of the environment itself. Terms involving these three potential energies are denoted together as $\mathcal{U}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda)$. U_{bias} is a biasing (umbrella) potential to facilitate sampling in the chemical coordinate λ. K_p ($p \in \{A, B, e, \lambda\}$) is the corresponding kinetic energy term for the conformational or chemical variable. From Eq. 1, one can derive coupled equations of motion for the atomic coordinates $\mathbf{r}_{A/B}$ and $\mathbf{R}_e$ and the chemical coordinate λ with a suitably assigned mass. Integrating the equations of motion subject to a holonomic constraint on $\lambda \in [0, 1]$ allows sampling of the "extended system" in a statistical ensemble of choice.[173] In the canonical ensemble, the partition function is

$$Z(\lambda, T) = \frac{\int \delta(\lambda - \lambda') e^{-\beta[\mathcal{U}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda') + U_{\text{bias}}(\lambda')]} d\mathbf{r}_A\, d\mathbf{r}_B\, d\mathbf{R}_e\, d\lambda'}{\int e^{-\beta[\mathcal{U}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda') + U_{\text{bias}}(\lambda')]}\, d\mathbf{r}_A\, d\mathbf{r}_B\, d\mathbf{R}_e\, d\lambda'} \tag{2}$$

where $\delta(\lambda - \lambda')$ is the Dirac-δ function and $\beta = 1/k_BT$ is the inverse temperature (k_B: Boltzmann constant, T: temperature). It follows that ΔG_{AB} is given by

$$e^{-\beta \Delta G_{AB}(T)} = \frac{Z(\lambda = 1, T)}{Z(\lambda = 0, T)} \tag{3}$$

Extension to multiple sites at one or multiple positions of a scaffold is generalized from the terms in $\mathcal{U}(\mathbf{r}_A, \mathbf{r}_B, \mathbf{R}_e, \lambda)$ above to

$$\begin{aligned}\mathcal{U}(\mathbf{r}_{\{s\}_{\{i\}}}, \mathbf{R}_e, \lambda) &= U_e(\mathbf{R}_e) + \sum_s^M \sum_i^{N_s} \lambda_{s_i} U(\mathbf{R}_e, \mathbf{r}_{s_i}) \\ &\quad + \sum_s^M \sum_{t>s}^M \sum_i^{N_s} \sum_j^{N_t} U(\mathbf{r}_{s_i}, \mathbf{r}_{t_j}) + U_{\text{bias}}(\{\lambda\}) \end{aligned} \tag{4}$$

Thus, each substituent i of the N_s substituents at each site s of the M total sites gets its own λ_{s_i}. Interactions of a substituent with itself and the environment $U(\mathbf{R}_e, \mathbf{r}_{s_i})$ are scaled by λ_{s_i} while interactions between sites $U(\mathbf{r}_{s_i}, \mathbf{r}_{t_j})$ are scaled by the product of $\lambda_{s_i}\lambda_{t_j}$, and all remaining interactions, $U_e(\mathbf{R}_e)$, are unscaled. Although λ-dynamics has been primarily implemented in CHARMM, it can be implemented in OpenMM using custom non-bonded forces.[8] However, for large chemical spaces the computational efficiency is poor. CpHMD methods based on λ-dynamics have also been implemented in GROMACS,[6] Amber,[5] and AMOEBA.[174]

In CHARMM, λ-dynamics is implemented through the BLOCK module, where many new features have been introduced to improve the accuracy, robustness, scope, and sampling. It is computationally expedient to ensure that λ remains between 0 and 1 (boundary constraint), and all λ values at a particular site add up to 1 (normalization). While these criteria can be maintained approximately or exactly with restraints or constraints, respectively, as was done in the earliest implementations,[173] it is more convenient to maintain them with implicit constraints through change of variables. This provides an alternative set of alchemical variables θ that map back to λ such that the constraints and normalization are satisfied by construction.[26]

Soft-core interactions that remove non-bonded singularities near the alchemical end points of 0 or 1 are important for convergence. They are especially critical for the accuracy and reproducibility of λ-dynamics because free energy is estimated by binning together states near alchemical end points, where hard cores can lead to very sharp changes in the free energy. The BLOCK module contains a special set of soft core functions for λ-dynamics that enables van der Waals and electrostatic interactions to be turned off concurrently.[175] For λ-dynamics, the PME electrostatics[35,176] gives better results than force switching electrostatics,[36] especially for longer simulations.[177,178] The BLOCK module includes commands to enable a generalization of PME for λ dynamics.[179] It is also worth noting that the MSLD command ('L' for λ) in the BLOCK module accepts an FFIX option that will run otherwise identical simulations, but with fixed values of λ for FEP validation or discrete λ sampling.[10,180−182] Several additions to BLOCK allow broader applicability of λ-dynamics to more unusual perturbations. Protein mutations to proline, and ligand calculations involving ring changes, core hopping, or macrocyclization require special considerations to ensure that when a substituent is non-interacting at $\lambda = 0$, the dummy atoms in the substituent are only bonded to one environment atom so they do not exert a net force on the rest of the system.[183] To satisfy these considerations, the BLOCK RMLA command allows removing λ scaling for classes of interactions, and it is recommended to only leave bond and angle interactions unscaled and to scale dihedrals. For finer granularity, soft bonds are implemented in BLOCK to, for example, break the proline ring at $\lambda = 0$ and allow free rotation of other amino acids at the same site around their λ backbone angle.[184,185] If significant portions of a molecule are similar but cannot be incorporated into the common core of a hybrid topology model due to differing charge or atom types, they may be harmonically restrained together with their bonded interactions scaled with the CATS command in BLOCK,[185] analogous to a similar process in NAMD.[7]

Another set of features crucial for sampling of the chemical space is adaptive landscape flattening (ALF) where a biasing potential in the λ space is iteratively developed to flatten the chemical landscape for enhanced sampling.[175,177,178] These biases are implemented by the LDIN and LDBV commands and are typically tuned by an external ALF python package.[175] Sampling can also be improved with Hamiltonian replica-exchange MD (REMD) through the REPD module.[27,31] More rapid sampling can be achieved with the BLaDE module[24] (Section 2.3).

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The above developments enabled sampling of massive chemical spaces spanning 512 HIV reverse transcriptase inhibitors,[27] 240 T4 lysozyme mutants,[177] and 32768 ribonuclease H variants,[171] as well as challenging perturbations of both ligands[186] and proteins.[185]

5.2. Hybrid Sampling and Free Energy Algorithms. The calculation of solvation free energy and binding affinity of small molecules to macromolecules are among the most important practical applications of MD simulations, especially with the potential impact on drug discovery efforts. A wide range of methodological advances were implemented in CHARMM to improve the statistical convergence and physical accuracy of free energy calculations. Conceptual advances in free energy methodologies implemented in CHARMM were reviewed in ref 187. For example, a version of λ-dynamics was introduced via a MC multicanonical REMD (FEP/REMD).[188,189] Specifically, the FEP/REMD helps resolve the poor convergence of the free energy estimates as a function of λ near the end points ($\lambda = 0$ and 1), which is often reflected as hysteresis between the forward ($0 \rightarrow 1$) and backward ($1 \rightarrow 0$) calculations from traditional FEP calculations based on single trajectories.

Applications to the calculation of the binding free energy of different kinase inhibitors demonstrated that the FEP/REMD algorithm was critical for tackling complex ligands accurately.[190−192] A similar general strategy improved the convergence of multidimensional US calculations by swapping configurations from different windows via Hamiltonian REMD (US/H-REMD)[193] (Section 7.1). Another issue concerns the sampling of solvent configurations. The binding of a ligand to a receptor frequently involves the displacement of a certain number of bound water molecules. This is not an issue if the binding site is in direct contact with the bulk solution. However, the convergence and accuracy in FEP/MD calculations can be severely compromised when a binding site is deeply buried and is inaccessible to bulk water. In this case, simple MD does not guarantee a complete sampling of the solvent during the FEP calculation. As an illustration, the binding of camphor to a deeply buried pocket in cytochrome P450cam causes about 7 water molecules to be expelled.[194] To address this, standard MD was coupled with the grand canonical MC (GCMC) algorithm to allow the number of water to fluctuate in any chosen region during an alchemical FEP calculation.[194] GCMC helps better sample the solvent configurations in the binding pocket that are poorly accessible to bulk solvent. It is also powerful by introducing fluctuations in the number of solvent molecules in FEP calculations carried out with a reduced model where only the region surrounding the binding site is explicitly considered while the effect of the surrounding solvent and protein is mimicked implicitly with the generalized solvent boundary potential (GSBP).[195] Such a strategy made it possible to calculate the standard binding free energy of antibiotics to the peptidyl-transferase P-site of the bacterial ribosome.[196,197]

Over the years, increasing efforts were made to streamline free energy calculations, enabling automated calculation of the absolute solvation free energy of a large number of small drug-like molecules using explicit solvent.[198] Moreover, collaborative efforts were made to test the accuracy and reproducibility of free energy calculations across different software packages.[199] One of the principal advantage of CHARMM is that different methodologies can be naturally integrated within a single job. For example, a US formulation of equilibrium binding[200] was used to characterize the binding specificity of a large number of SH2 domains[201] with the generalized Born with a simple switching (GBSW) implicit solvent model.[131] As another example, the PBEQ continuum electrostatics module of CHARMM[108,202] conveniently allows one to directly access and read MD trajectory snapshots, and then combine its MM potential energy together with the solvation contribution based on the Poisson−Boltzmann and surface area approximation (PBSA). This MM/PBSA strategy, seamlessly integrated within CHARMM, has been used, for example, to process a large number of protein complexes to assess the binding specificity within a family of synaptic surface receptors.[203]

A growing family of hybrid sampling methods combining the strength of MD and Metropolis MC were tested and implemented, benefiting from the flexibility of the control flow from the native CHARMM scripting command language at the level of the input file.[204,205] These algorithms typically consider new configurations generated by driving the system via a non-equilibrium MD (NEMD) trajectory that are subsequently treated as putative candidates for MC acceptance or rejection.[204,205] The hybrid NEMD/MC algorithms can be exploited in a variety of context and offer a promising avenue to sample the configurations of complex systems. For example, the discrete ionization state of titratable residues can be sampled, effectively as a constant-pH simulation.[206] Another example is to consider new configurations of an all-atom system generated by driving it via NEMD toward a configuration that originated from a CG simulation. It was shown that the CG-guided hybrid NEMD/MC algorithm can enhance the sampling of solvated peptides even with fairly rudimentary CG models as a guide.[207]

5.3. Optimal Variance Alchemical Path for Free Energy Calculation. Despite continuous development of free energy calculation methods,[28,99,208] practical challenges impede their precision and possibly reliability.[45,199] Options for improvement include enhanced sampling,[99,189,209−211] careful design of alchemical cycles,[199,211,212] variational and integration approaches,[211,213−216] and the design of the alchemical path itself,[99,215,217−220] the latter being the focus of this section.

The hybrid Hamiltonian method relies on the 'optimal alchemical path' theory.[217] To overcome barriers between reactant and product phase spaces, it is implemented at the interaction pair level, treating each pair separately though in parallel. Denoting abolished (A) interacting pairs as $p_i \in P_A$ and created (B) ones as $p_i \in P_B$, the corresponding Hamiltonian contributions are

$$H_{\mathrm{Oab}}(p_i, \lambda) = -\frac{2}{\beta}\ln([1-\lambda] + \lambda e^{-\beta(H_A(p_i)-G_{A0,i})/2})$$
$$H_{\mathrm{Ocr}}(p_i, \lambda) = -\frac{2}{\beta}\ln([1-\lambda]e^{-\beta(H_B(p_i)-G_{B0,i})/2} + \lambda) \quad (5)$$

where H_A, $G_{A0,i}$, H_B, and $G_{B0,i}$ are the energy functions and estimates for the free energy of abolishment or creation of each of the pairs, p_i. Approximations for Eq. 5, denoted by H_{cr} and H_{ab}, follow from Eq. 25 of ref 217, and the hybrid Hamiltonian is given by

$$H_{HH}(\lambda) = \sum_{p_i \in P_A} H_{\mathrm{ab}}(p_i; \lambda) + \sum_{p_i \in P_B} H_{\mathrm{cr}}(p_i; \lambda) + H_C(\text{terms} \notin P_A \cup P_B) \quad (6)$$

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

where H_C is for all other terms unaffected by the transformation. The improper and proper dihedral angle fluctuations being modest, simple multiplication factors are used for created and abolished terms, respectively (isomorphous to Eq. 25 in ref 217). For Ewald sum, a linear scheme for charge, $q = q_C + (1 - \lambda)q_A + \lambda q_B$, is used.

As a result, the derivative with respect to λ can be intertwined as an additional dimension to that of the system spatial coordinates, **r**, extending Eqs. 4.6 and 4.9 of ref 176 as

$$\frac{\partial Ew\_Sum}{\partial\{\mathbf{r}, \lambda\}} = \left\{\frac{\partial Q(q)}{\partial\{\mathbf{r}\}}, Q\left\{\frac{\mathrm{d}q}{\mathrm{d}\lambda}\right\}\right\}\cdot(\theta_{rec}{}^{*}Q(q)) \tag{7}$$

where $*$ indicates convolution, Q the charge mesh, and θ_{rec} the reciprocal factor mesh defined in ref 176. An application of the method was on the R67 DHFR system that is a pseudo-homotetramer, a dimer of dimers.[221−223] To simulate the mutation process, the two subunits of one dimer had an *A*-hybrid residue at position 59 and the two subunits of the other dimer had a *B*-hybrid residue at position 62 (Figure 2).

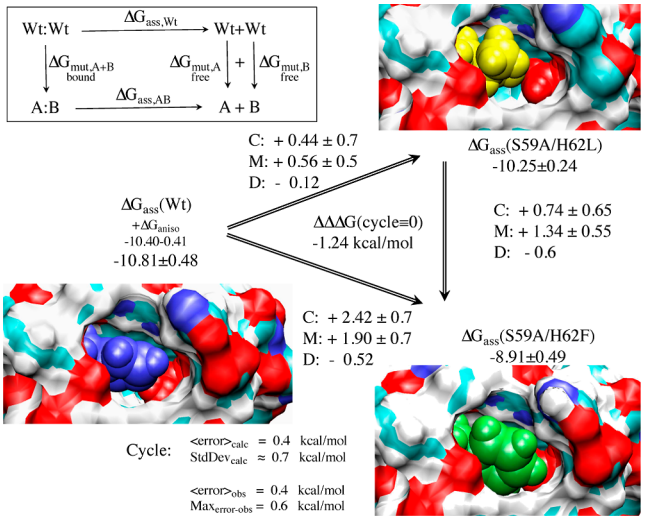


Figure 2. Thermodynamic cycle. Inset: general cycle design; horizontal arrows: measured affinities in kcal/mol;[222,223] vertical arrows: computations for tetramer (left) and the two types of dimers (right). Graphical panels: local molecular surface at the interfaces with the mutated residues displayed as spheres for WT (S59: red, H62: blue), S59A/H62L (A59: red, L62: yellow), and S59A/H62F (A59: red F62: green). For WT, a 0.41-kcal/mol entropic term is added to account for higher symmetry.[222] Computed (C:), measured (M:) differences, and discrepancies (D:) are given. Global discrepancy ($\Delta\Delta\Delta G$) for the 3 cycles provides a self-consistency check. Average standard deviation ($StdDev_{calc}$) and error ($\langle error\rangle_{calc}$) were computed using autocorrelation functions[224] considering λ windows as independent. The average error $\langle error\rangle_{obs}$ and the maximum observed error $Max_{error\text{-}obs}$ that compare experimental results with calculations are also reported.

Simulations were run sequentially for 10 discretized values of λ from 0 to 1. Hybrid residues were also subjected to TI in their isolated acetylated and aminated form as a control.

Branches of the various thermodynamic cycles in Figure 2 are further analyzed in Figure 3. Individual curves are bell-shaped, mirroring the quadratic form of the partition function of the optimal path as function of λ (Eq. 11 of ref 217) and yielding a linear integrand. Due to differences in the position of the maxima for the different branches of a same cycle, the

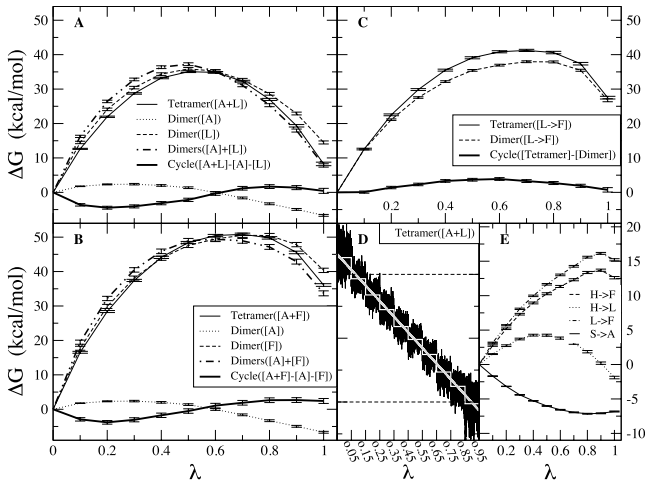


Figure 3. Integration along the thermodynamic cycles in Figure 2. Cumulative error estimates[224] are also shown for the tetramer, each of the dimers, the sum of the dimers, and the global cycle for (A) WT to S59A/H62L, (B) WT to S59A/H62F, and (C) S59A/H62L to S59A/H62F for which only one dimer is involved since residue 59 remains as Ala. (D) Integrand for the transformation of the tetramer from WT to S59A/H62L shown in black as an example. The average for each λ window is marked by a stepwise white line. Linear regression along the whole trace is shown as a light gray line to appraise the linearity of the integrand with respect to λ. Dashed lines mark ±100 kcal/mol. (E) Integrand for isolated hybrid residues (acetylated and aminated) in a vacuum to evaluate the intrinsic energy contributions due to the FF energy difference of the original residues.

global cycle profiles are sinusoidal rather than quadratic, nonetheless very tempered.

As previously reported, S59A/H62L is favorable despite loss of a hydrogen bond and the formation of a small hydrophobic cavity.[222] In comparison, S59A/H62F is less favorable despite good shape complementarity and creation of new hydrophobic contacts.[223] Despite modest calculation effort only intended to illustrate the optimal alchemical path integrand properties, simulations reproduce those unexpected results. Interestingly, TI on isolated hybrid residues revealed the predominance of the amino acids intrinsic FF potential differences on the integrand, suggesting that reducing those differences could reduce the difficulty to reach accurate results. The linearity of the integrand with respect to λ for the method presented here facilitates integration, hence it is a desired property. It avoids the need for evolved integration schemes that can amplify errors, but are required to treat irregularity or singularity found, for example in conventional van der Waals creation.[214]

6. CONSTANT PH METHODS

6.1. Hybrid-Solvent and All-Atom Continuous Constant pH Methods. Describing protonation state changes due to a change in solution pH or conformational environment was first enabled in CHARMM through the GB CpHMD methods.[225,226] In these methods, an auxiliary set of (λ) coordinates representing the evolution of protonation states are propagated based on the idea of λ-dynamics[173] (Section 5.1). Since 2010, the CpHMD framework was further developed to be carried out in explicit-solvent MD simulations (see, e.g., Section 5.1). An example is the hybrid-solvent CpHMD[227] that samples solute conformation in explicit solvent but leverages the GBSW implicit solvent model[131] for propagating protonation states. The pH REMD method

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

was also developed to accelerate sampling of the coupled conformation and protonation states.[227] The hybrid-solvent CpHMD was later extended for transmembrane protein simulations[228] by including the implicit membrane GBSW model[229] with a water cylinder to account for water molecules in the pore of a channel or a transporter. To remove the dependence on the GB models which limits the accuracy, the all-atom CpHMD methods with generalized reaction field[230] or PME for long-range electrostatics[179] have been developed. To enforce net charge neutrality in all-atom CpHMD, an approach based on cotitrating ions[230] or water[231] has been developed. The hybrid-solvent and all-atom CpHMD have enabled not only new lines of inquiries, e.g., pH-dependent self-assembly mechanism of chitosan in which a total of 160 glucosamine units were allowed to titrate,[232] but they also provided fresh perspectives to resolve old questions where, e.g., the hybrid-solvent CpHMD simulations revealed the formation of proton-coupled hydrogen bonds as a major determinant for acid/base.[233]

6.2. Constant pH MD with Discrete Protonation States. There are two main classes of constant pH simulations depending on whether the protonation states vary discretely (either deprotonated or protonated)[207,234−241] or continuously.[172,225−227,242] Two types of the former class are implemented in CHARMM. The first is based on the MD/MC constant pH method[237,238] that is available only in implicit solvent. The second type is based on the EDS method,[37] and is available for explicit solvent. Constant pH simulations with continuous protonation states, also in CHARMM, are available for implicit,[225] combined implicit and explicit,[227] as well as explicit-only solvent.[172]

The MD/MC method,[237] originally available for Amber, has been implemented in CHARMM and further extended to include constant pH REMD.[238] During MD simulation, attempts to change the protonation state according to the Metropolis criterion are made at a user-defined interval. The deprotonated state is modeled with the proton present following the charge distribution for the deprotonated state. For replica i, let the positions and momenta of atoms be q_i and p_i, respectively, N_i^p be the number of titratable residues that are protonated, and pH_l its pH. Similarly define q_j, p_j, N_j^p and pH_m for replica j. Denoting $X_i^l \equiv (q_i, p_i, N_i^p, \mathrm{pH}_l)$ and $X_j^m \equiv (q_j, p_j, N_j^p, \mathrm{pH}_m)$, the probability of exchange between replica i and replica j is

$$w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = \begin{cases} 1 & (\Delta \le 0) \\ e^{-\Delta} & (\Delta > 0) \end{cases} \tag{8}$$

$$\Delta \equiv \ln 10 \cdot (\mathrm{pH}_m - \mathrm{pH}_l)(N_i^p - N_j^p)$$

In addition to the constant pH REMD that greatly improves sampling of the protonation state,[238] the reservoir constant pH REMD method was developed to better sample conformational states.[243] It relies on pregenerated reservoirs of conformations with fixed protonation states. The reservoirs can be generated either by long MD simulations, or with an enhanced sampling method, so that conformations with a given protonation state follow the Boltzmann distribution. Then an attempt to replace the current conformation with a random reservoir structure is made after a given number of steps. The attempt is accepted if protonation states of all ionizable residues match with those of the reservoir structure, and rejected if not. In this way, the system can sample conformations from the Boltzmann ensemble of the reservoir.

Another method implemented in CHARMM is the EDS with Hamiltonian REMD (EDS-HREM).[40] In the EDS approach,[37] a hybrid Hamiltonian enveloping both states is defined such that the corresponding partition function is the sum of partition functions for individual Hamiltonians. In addition, a smoothness parameter can be introduced to facilitate conformational transitions between states with high energy barrier. In its constant pH implementation,[40] the two states are protonated and deprotonated, and a pH-dependent energy offset between the two states is introduced.

$$E_{\mathrm{EDS}}(\mathbf{x}, s, \mathrm{pH}) = -\frac{1}{\beta s} \ln\left(\sum_{i=1}^{N} e^{-\beta s [E_i(\mathbf{x}) - E_i^{\mathrm{offset}}(\mathrm{pH})]} \right) \tag{9}$$

Here, $E_i(\mathbf{x})$ is the potential energy of state i with coordinate $\mathbf{x}$, s is the smoothness parameter, and $E^{\mathrm{offset}}(\mathrm{pH})$ is the pH-dependent energy offset calculated ahead of the simulation via thermodynamic cycling. The Hamiltonian can be extended to several titrating groups, conveniently describing clusters of coupled residues. For example, it has been used to calculate the pK_a values of four glutamic acid residues in the selectivity filter of a sodium channel.[244]

Different replicas have different values of s, which allows for replicas with low s (very smoothed) to cross energy barriers, while replica with $s = 1$ yields the conformational ensemble identical to the semigrand canonical ensemble at convergence. As a follow-up, a 2-dimensional (2D) replica exchange pH method was added in CHARMM, where the second dimension is pH.[41]

FEP methods have also been used for protein pK_a calculation in both implicit and explicit solvent,[245] as well as in QM/MM settings.[246] In a recent study, pK_a calculations from the 2D EDS-HREM in explicit solvent have been found to agree well with FEP results for a complicated system consisting of four selectivity filter glutamate residues of an ion channel with bound ions.[247] Additional FEP simulations led to a new proposed mechanism of selectivity in this ion channel, based on the shift of the pK_a value in the presence of different ions.[247]

6.3. Proton Hopping Simulations. Classical biomolecular MD simulations normally do not allow changes in covalent bonding. This is an issue in systems involving proton transfer, as e.g. in proton diffusion in water where a proton breaks a bond with one water and forms a new one with a neighboring water molecule. The MOBHY (for "mobile hydrogen") module in CHARMM allows proton mobility by interspersing discrete proton moves during a dynamics trajectory.[248] After a given number of MD steps, an attempt is made to move a titratable proton to an eligible alternative location, i.e., a potential acceptor to which the titratable proton is hydrogen bonded. Upon the hop attempt, the molecular geometries and FF parameters of protonated and deprotonated species are changed accordingly. The missing protons are represented by dummy atoms (no charge and no interactions with surroundings). The initial protein structure is generated with all potential protons present, i.e., all specified titratable residues should be fully protonated in the PSF (whether they are truly protonated is selected by the user). Thus, no actual changes in bonding take place during a proton hopping simulation; only the atom types and charges change. The excess proton is represented as a classical hydronium ion. Acceptance of a

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

proton move is based on a Metropolis-like criterion that employs an empirical threshold for the energy change upon proton hopping. The threshold is chosen to reproduce the experimental proton diffusion coefficient in water. Similar empirical thresholds are used for proton hopping between water and protein side chains, while the true rates can be obtained by more elaborate methods.[249] This method has been applied to proton conduction by gramicidin A,[248] investigation of the asymmetry of proton conduction in the influenza M2 proton channel,[250] and evaluation of models for the human voltage gated proton channel.[251]

7. ENHANCED SAMPLING AND TRANSITION PATH METHODS

7.1. Replica Exchange MD (REMD). In REMD, N independent copies (or replicas) of a system are run in parallel and are periodically swapped (i.e., exchanged) to enhance the crossing of potential energy barriers.[252,253] REMD is useful in systems where energy barriers lead to poor sampling and slow convergence in conventional MD, hindering accurate calculation of thermodynamic quantities.[188,189,254]

When the system volume does not change, the probability of observing a system in a configuration represented by coordinates X and Hamiltonian a with energy $E_a \equiv E(X)$ is

$$P(X, E_a) = \frac{e^{-E(X)\beta_a}}{Z_a} \tag{10}$$

where $\beta_a = 1/k_B T_a$, with Boltzmann constant k_B and temperature T_a. Z_a is the partition function. For N non-interacting replicas, the probability of observing the system in a particular state is the product of the probabilities for individual replicas:

$$P = \prod_{n=1}^{N} P_n \tag{11}$$

The enhancement of sampling in REMD comes from periodic swapping of the coordinates and velocities between two replicas. By imposing detailed balance, the ratio of the forward and backward transition rates between replicas a and b in exchanging their coordinates X and Y is given by[253]

$$\frac{P_{\text{forward}}}{P_{\text{backward}}} = e^{(\beta_b - \beta_a)(E(Y) - E(X))} \equiv e^{\Delta_T} \tag{12}$$

where e^{Δ_T} is the temperature replica exchange probability. In Hamiltonian REMD, e^{Δ_H} can be similarly defined. They are used to accept or reject the exchange using the Metropolis criterion:

$$P_{\text{exchange}} = \begin{cases} 1 & (\Delta_{\{H,T\}} \geq 0) \\ e^{\Delta_{\{H,T\}}} & (\Delta_{\{H,T\}} < 0) \end{cases} \tag{13}$$

When the system's volume changes (NPT ensemble), Eq. 10 changes to[255]

$$P(X, E_a) = \frac{e^{-\beta_a[E(X) - P_a V(X)]}}{Z_a} \tag{14}$$

where P_a is the external pressure at Hamiltonian a and $V(X)$ is the volume of the coordinates X. The exponents of P_{exchange} then become

$$\Delta_{\{H,T\},P} = \Delta_{\{H,T\}} + \Delta_{\text{correction}}$$
$$\Delta_{\text{correction}} = (\beta_b P_b - \beta_a P_a)(V(Y) - V(X)) \tag{15}$$

REMD in CHARMM is handled through the REPD (REPlica Distributed) command. It requires MPI parallelism with one or more MPI processes per replica. Exchanges are attempted at a user-specified interval, typically on the order of 1 ps. The exchange direction alternates between "up" and "down" in the replica space. While it is not strictly necessary to attempt exchanges only between neighboring replicas, acceptance of an exchange between two replicas requires overlap between their potential energy distributions that is typically highest for neighboring replicas. Exchanges are accomplished by swapping coordinates and velocities between MPI processes, so that each MPI process yields a "replica" trajectory, i.e., coordinate frames corresponding to a single temperature or Hamiltonian.

CHARMM supports REMD for temperature, general Hamiltonian, self-guided Langevin,[256] and CpHMD using either discrete[237,238] or continuous[179,225,227] protonation states. CHARMM also supports coupling of the top and/or bottom replicas (i.e., the highest and lowest in replica space) to pregenerated structure reservoirs. Exchanges with the reservoir can further accelerate conformational sampling,[257] and can be done assuming either Boltzmann (recommended)[258] or non-Boltzmann[259] weighting. In constant pH REMD, CHARMM supports exchanges with reservoirs that have fixed protonation states, where exchanges with a structure in the reservoir can only be accepted if the protonation state of all ionizable residues matches the structure to be exchanged (Section 6.2).[243]

CHARMM also supports multidimensional REMD,[260,261] with the only restriction being that a dimension aside from the general Hamiltonian may only be used once (e.g., one temperature dimension and one self-guided Langevin dimension is permitted, but not two temperature dimensions). The combination is multiplicative: For example, a setup with 4 temperatures and 2 Hamiltonians will use 8 replicas in total. To simplify scripting, CHARMM sets up several user-accessible variables, such as ?NREP and ?MYREP, which refer to the total number of replicas and the global replica index respectively. For multidimensional REMD, ?NREPD⟨X⟩ and ?MYREPD⟨X⟩ refer to the total number of replicas and replica index in dimension ⟨X⟩, respectively (for example, ?NREPD1 is the number of replicas in the first replica dimension).

REMD in CHARMM can be combined with other ensemble methods such as EDS[37] via the MSCALE module.[262] Earlier, a constant pH method in explicit solvent with discrete protonation states was developed based on a combination of EDS and a 1D REMD.[40] A more recent version features EDS with a 2D REMD (the second dimension being pH), which significantly accelerates the convergence of constant pH simulations.[41]

7.2. Biasing Methods. *7.2.1. Targeted MD (TMD).* Conformational transition pathways can be simulated with a number of TMD methods. The original implementation[263] introduces a holonomic constraint that reduces the RMSD from the target coordinates with a preset value at each MD step. While this guarantees to reach the target conformation, generated pathways are generally irreversible[264] and they can cross large free energy barriers.[265] By using a perturbation of a

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

fixed magnitude that minimizes the RMSD with the target at every step, the restricted perturbation TMD (RPTMD) method[265] generates low free energy pathways along which potential of mean force (PMF) profiles can be readily calculated.[266] The RMSD can also be decreased by a restraint potential (RTMD) that can be symmetrized to yield more reversible paths.[264] Due to the use of global best-fit rotations, these TMD methods tend to favor large scale motion before small conformational changes,[265,267] which is subdued in locally restrained TMD (LRTMD) by applying a number of TMD restraints on subsets of atoms.[267]

7.2.2. Related Conformational Free Energy Sampling. CHARMM supports a number of enhanced sampling techniques to evaluate conformational free energy differences. US[138] and adaptive US[268,269] of distances, angles, torsions, RMSD, and more complex geometrical order parameters are supported by the CONS, RXNCOR, and ADUMB modules. US is typically performed through the use of harmonic restraints that bias the system toward a desired target. CHARMM also supports best-fit positional restraints in which the reference coordinates are first rotated and translated to minimize the restraint energy. These best-fit restraints are key to the efficiency of confinement methods[270−274] that calculate conformational free energy differences by transforming (part of) the system to the desolvated harmonic oscillator state. The Gaussian-mixture US (GAMUS) method allows enhanced sampling of multidimensional order parameters (3−6 dimensions).[275,276] Like adaptive US, GAMUS uses the negative of the calculated free energy as the biasing potential, which is updated periodically while taking all sampled data into account. GAMUS constructs its biasing potential from a Gaussian-mixture model that fits the probability distribution using fully optimized Gaussian functions. By foregoing grids, GAMUS can sample higher dimensional spaces than traditional adaptive US. CHARMM also supports Tsallis-based biasing potentials[277] that increase sampling by reducing the force near energy barriers. In CHARMM, Tsallis-based sampling can also be coupled to replica exchange with solute tempering[278,279] for faster sampling.[280,281]

7.3. String Method (SM) for Conformational Transitions. If a process of a system with positions $\mathbf{x}$ is described by the reaction coordinate $q(\mathbf{x}) \in \mathbb{R}$, the free energy $\mathcal{F}$ of a conformational state $q(\mathbf{x}) = q_0$ is

$$e^{-\beta\mathcal{F}(q_0)} = \int e^{-\beta E}\delta(q(\mathbf{x}) - q_0)\, d\mathbf{x} \tag{16}$$

One often wishes to follow the progress of an actual chemical or physical reaction as q_0 is varied from the initial (reactants) to the final (products) value. Below, we focus on a set of methods in which the reaction coordinate is optimized from an initial pathway or a set of intermediate configurations.[282−287]

The essential idea of SM[288−290] is to assume that the optimized path is everywhere tangent (possibly up to a constant multiplicative tensor) to the reaction coordinate gradient without needing to specify an analytical form for it (Figure 4). Three versions of SMs implemented in CHARMM are described below.

7.3.1. Zero-Temperature SM. The zero-temperature SM (ZTSM) computes a minimal-energy path (MEP) which is a curve in the space of N_a atom coordinates defined as $C = \{\mathbf{x}(\alpha) \in \mathbb{R}^{3\times N_a}, \alpha \in [0, 1]\}$ that, for any $\alpha \in (0, 1)$, satisfies

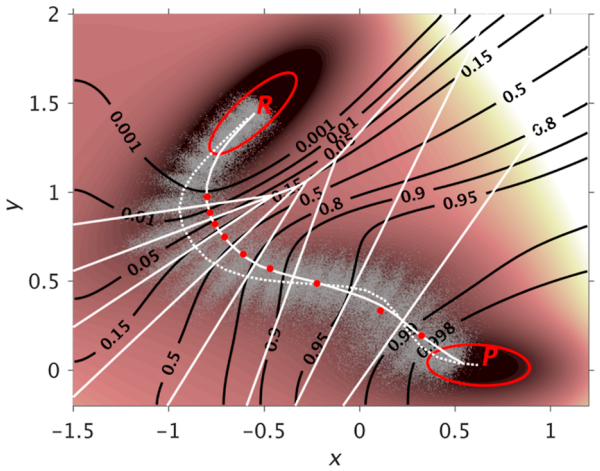


Figure 4. Illustration of the SM on the 2D Mueller potential. An MEP (dotted white curve) and a finite temperature string (solid white curve) connect the reactant (R; $q = 0$) and product states (P; $q = 1$) enclosed within red ellipses. Black contours represent isocommittor surfaces obtained from a 2nd order finite difference solution of the backward Kolmogorov equation for overdamped Langevin dynamics. White straight lines are planar approximations to the isocommittor surfaces, which also partition the configurational space into a Voronoi tessellation with nodes (red bullets). Gray dots are simulation coordinates from overdamped Langevin dynamics restrained to reaction coordinate planes and collectively define a transition tube.

$$\mathbf{x}'(\alpha) = \frac{d\mathbf{x}}{d\alpha} \| \nabla_{\mathbf{x}} E(\mathbf{x}) \tag{17}$$

where $\mathbf{x}(0)$ and $\mathbf{x}(1)$ correspond to the reactant and product state, respectively. The ZTSM evolves an initially assigned guess to the MEP using the steepest descent (SD) minimization while enforcing uniform parametrization by arc length, $|d\mathbf{x}/d\alpha|$ = constant. In the CHARMM implementation, the continuous string is discretized into N replicas or 'images,' each assigned to a separate group of processors for parallel execution

$$\mathbf{x}_i = \mathbf{x}(\alpha_i),\ i = 1, \ldots, N \tag{18}$$

with N typically determined by the available computing processors. The string evolves to the MEP as

$$\hat{\mathbf{x}}_i(\tau + \Delta\tau) = \mathbf{x}_i(\tau) - \Delta\tau\gamma_0^{-1}\nabla_{\mathbf{x}} E(\mathbf{x}_i) \tag{19}$$

$$\mathbf{x}_i(\tau + \Delta\tau) = [R\hat{\mathbf{x}}(\tau + \Delta\tau)]_i \tag{20}$$

where $\Delta\tau\gamma_0^{-1}$ controls the speed of SD evolution ($\Delta\tau$ is an artificial time step, and γ_0 is a friction constant that ensures dimensional consistency).

Eq. 19 is advanced independently for each image, and R is the reparameterization operator that corrects the provisional coordinates $\hat{\mathbf{x}}$ so that $|\mathbf{x}(\alpha)'|$ is constant along the string. R is common to the SMs in CHARMM (Section 7.3.4). The evolution step ($\Delta\tau\gamma_0^{-1}$) and convergence criteria can be set manually, or automatically by the SD minimizer of CHARMM. While SD is the default minimizer for ZTSM, other minimizers in CHARMM can also be used.

7.3.2. Finite-Temperature SM. The finite-temperature SM (FTSM) can be derived from the backward Kolmogorov equation (BKE)[291] corresponding to overdamped Langevin dynamics.[292−295] In FTSM, the desired reaction coordinate q is assumed to be the *committor* function that solves the BKE,[291]

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

and the committor isosurface $q = q_0$ is approximated by a hyperplane (see Figure 4)

$$|\nabla q(\mathbf{x})|\delta(q(\mathbf{x}) - q_0) = \delta(\boldsymbol{\nu}(q_0)\cdot[\mathbf{x} - \boldsymbol{\phi}(q_0)]) \tag{21}$$

The Jacobian $|\nabla q(\mathbf{x})|$ preserves the volume and $\boldsymbol{\nu}(q_0)$ is the unit normal to the hyperplane P_{q_0} that approximates the isosurface $q(x) = q_0$. $\boldsymbol{\phi}$ is constrained by

$$\begin{aligned}\boldsymbol{\phi}(q_0) &= Z(q_0)^{-1}\int \mathbf{x}e^{-\beta E}\delta(\boldsymbol{\nu}(q_0)\cdot[\mathbf{x} - \boldsymbol{\phi}(q_0)])\,d\mathbf{x}\\ &= \langle\mathbf{x}\rangle_{P_{q_0}}\end{aligned} \tag{22}$$

where $Z(q_0) = \int e^{-\beta E}\delta(\boldsymbol{\nu}(q_0)\cdot[\mathbf{x} - \boldsymbol{\phi}(q_0)])\,d\mathbf{x}$ is the partition function of the hyperplane. In analogy with an MEP, we can parametrize a continuous curve $\boldsymbol{\phi}(q_0(\alpha))$ having $dq_0/d\alpha > 0$, and identify it with the average reaction path. Provided that the transition 'tube' (Figure 4), as measured by the variance of $|\mathbf{x} - \boldsymbol{\phi}(q_0)|$, is not too large, $\boldsymbol{\phi}$ also represents the dominant reaction path. From Eqs. 21 and 22, it can be shown that[292]

$$\boldsymbol{\nu}(q_0)\|\frac{d\boldsymbol{\phi}(q_0)}{d\alpha} \tag{23}$$

i.e., the reaction coordinate hyperplanes are locally perpendicular to the reaction path (string). In FTSM, Eqs. 22 and 23 are iteratively solved.[285,293,294] From an approximation to the string at iteration n ($\boldsymbol{\phi}^n$), one obtains $\boldsymbol{\nu}^n$ using Eq. 23, which permits computing $\boldsymbol{\phi}^{n+1}$ using Eq. 22. This is repeated until $\boldsymbol{\phi}^n$ does not change (up to thermal noise). The free energy can then be obtained by TI of the free energy derivatives sampled on the hyperplanes,[285,294] or by sampling a Voronoi tessellation (Figure 4).[290]

The FTSM in CHARMM can optionally use Hamiltonian REMD to accelerate sampling, and an upper bound on the transition tube width can be set to limit sampling near a predefined path. In a parallel implementation,[296] FTSM starts from an initial string discretized into N images $\boldsymbol{\phi}_i^0$, $i \in \{1, ..., N\}$, which can be obtained from, e.g., an MEP or a biased dynamics trajectory. To each image $\boldsymbol{\phi}_i$ one assigns a separate CPU group and a complete all-atom MD simulation system denoted by $\mathbf{x}_i$, to be used for sampling each reaction coordinate hypersurface. Each CPU group receives the neighbor images $\boldsymbol{\phi}_{i\pm1}$ in addition to ϕ_i, which are required to compute $\boldsymbol{\nu}(q_0)$ in Eq. 23, and samples the hyperplanes independently of the other groups.

7.3.3. String in Collective Variables. There are cases when variables other than Cartesian coordinates, e.g., distances,[297] are more suitable for the reaction coordinate. Following the steps in ref 289, SM in CHARMM has been reformulated in a coarse-grained (CG) space of collective variables (CVs).[139] Assume that the reaction coordinate is determined by a set of CVs $\theta_j(\mathbf{x})$ ($j = 1, ..., K$) *via* some function f (which does not need to be specified explicitly): $q(\mathbf{x}) = f(\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), ..., \theta_K(\mathbf{x}))$. The coarse-graining leads to a K-dimensional free energy landscape as a function of CV coordinates denoted by $\mathbf{z}$:

$$\begin{aligned}F(\mathbf{z}) &= -\frac{1}{\beta}\ln\left(\frac{1}{Z}\int_{\mathbb{R}^{3\times N_a}} e^{-\beta E(\mathbf{x})}\prod_{j=1}^{K}\delta(z_j - \theta_j(\mathbf{x}))\right)\\ &= -\frac{1}{\beta}\ln\langle\delta(\mathbf{z} - \boldsymbol{\theta}(\mathbf{x}))\rangle\end{aligned} \tag{24}$$

and a metric tensor $\mathbf{M}(\mathbf{z})$ defined by

$$\frac{M_{jk}(\mathbf{z})}{e^{\beta F(\mathbf{z})}} = \sum_{l=1}^{N_a}\frac{1}{m_l}\langle\nabla_{\mathbf{x}_l}\theta_j(\mathbf{x})\cdot\nabla_{\mathbf{x}_l}\theta_k(\mathbf{x})\delta(\mathbf{z} - \boldsymbol{\theta}(\mathbf{x}))\rangle \tag{25}$$

where m_l is the mass of atom l. Using F and $\mathbf{M}$, it is possible to write down Langevin equations governing the evolution of $\mathbf{z}$.[289] Further, assume that the reaction proceeds *via* a localized reaction channel that contains a minimum *free* energy pathway (MFEP) on the CV landscape

$$\mathbf{z}'(\alpha)\|\mathbf{M}\nabla_{\mathbf{z}}F \tag{26}$$

with parameter $\alpha \in [0, 1]$ and $|\mathbf{z}'(\alpha)|' = 0$ (equal arc length) in analogy with Eq. 17 for the MEP.

The SM in collective variables is an iterative algorithm for computing the MFEP using local averaging of the force $\nabla_{\mathbf{z}}F$ and metric tensor $\mathbf{M}$ obtained from restrained MD simulations.[139,289] After the string converges to the MFEP, two types of free energy profiles can be computed, $F[\mathbf{z}(\alpha)]$ in the K-dimensional space of the CVs, and a 1-dimensional profile $\mathcal{F}(\alpha)$ associated with the reaction coordinate hyperplanes on $\mathbf{x}$. An approximate calculation of $\mathcal{F}(\alpha)$ in CHARMM is implemented using Voronoi tessellation, which also allows computation of the mean first passage time along the reaction coordinate using the Markov state model.[290]

7.3.4. Reparameterization. The SMs described here involve optimization of continuous curves (strings) specified by a parameter, e.g., $\{\boldsymbol{\phi}(\alpha),\ \alpha \in [0, 1]\}$. In numerical implementation, a set of discrete points along a string are used instead. To maintain uniform string resolution, parametrization by arc length is used, i.e.,

$$\alpha = \int_0^{\alpha}|\boldsymbol{\phi}'|\,d\alpha\Big/\int_0^1|\boldsymbol{\phi}'|\,d\alpha \tag{27}$$

which implies that $|\boldsymbol{\phi}'|$ is constant along the string, or that $|\boldsymbol{\phi}_i - \boldsymbol{\phi}_{i-1}|$ is constant for all images $i > 0$. Because the string deforms as it evolves, points (images) along the curve $\phi(\alpha)$ must be periodically reassigned to satisfy equidistance. This *reparameterization* operation (R in Eq. 20), is implemented by interpolating the string onto a refined parameter grid, i.e., α_j, $j = 1, ..., N_f$, with $N_f = 5 \times N$, computing arc length on this grid normalized to the unit interval, and interpolating onto the original uniform parameter grid. Linear interpolation is the default and recommended method. Others such as B-splines and cubic splines can also be used.

7.3.5. String with Swarms-of-Trajectories. Rather than refining the string in the multidimensional space of CVs by estimating the average force and metric tensor from restrained trajectories via Eqs. 24 and 25 as described above, an alternative approach considers the average dynamic drift of those variables determined on-the-fly via ensemble of short unbiased trajectories starting at different points along the string.[298] One advantage of this so-called "SM with swarms-of-trajectories" over the traditional procedure is that the computational task can be naturally distributed over many computer nodes with negligible interprocessor communication. The formal equivalence between the two approaches in the limit of very short trajectories was established,[299,300] and their respective significance has been clarified.[300]

7.3.6. Script-Based SM Approach and Structure Building. In a first application of the SM with swarms-of-trajectories to an all-atom solvated protein,[298] the activation pathway of Hck kinase and the inactivating DFG-flip were determined.[301,302] It

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

bears emphasizing that the SM could be scripted directly in the input file of CHARMM, and required no new source code. The powerful scripting facilities within CHARMM, especially the ability to modify the bonding topology of the system on the fly using the Patch Residue (PRES) facility, made it possible to generate all-atom models of the polymerized FT-30 membranes, which are widely used in reverse osmosis operations.[303,304]

7.4. Adaptively Biased Path Optimization (ABPO) for Transition Path Sampling. Algorithms to compute the energetics and conformations associated with protein conformational transitions are most often based on path-restrained sampling using a chain-of-states defined at specified intervals along the path. The ABPO method[305] is an alternative approach that does not require the protein system be restrained to the path. ABPO is implemented in CHARMM through the ENSEMBLE module with options for defining CVs (also called reduced variables, RVs) and path optimization parameters. An adaptive biasing potential, V_b, is utilized to enhance sampling of the path without restraining the system to specific points on the path[306]

$$V_b(\lambda, t) = k_B T \frac{b}{1-b} \ln[c(1-b)h(\lambda, t) + 1] \tag{28}$$

where b is the fraction of the free energy flattened by the bias, c has an inverse time unit and controls how the bias couples to the dynamics. V_b adapts from the sampling histograms $h(\lambda, t)$ that counts visits to the region of the path around λ over time t. The PMF is a direct result of the adaptive bias potential obtained for the optimal path.

A second distinction of ABPO compared to path-restrained methods is that evolution of the ABPO path begins by initiating multiple trajectories from an equilibrium ensemble simulated at each end state (Figure 5A). As such, the generation of unphysical structures at specified intervals along the initial chain-of-states path is avoided when starting the ABPO calculation. As trajectories move out of the end-state basin, their proximity to the path is retained with a tube potential of specified radius and centered on the path. An advantage of free sampling within the tube is to reduce frustration in sampling a rugged free energy landscape.

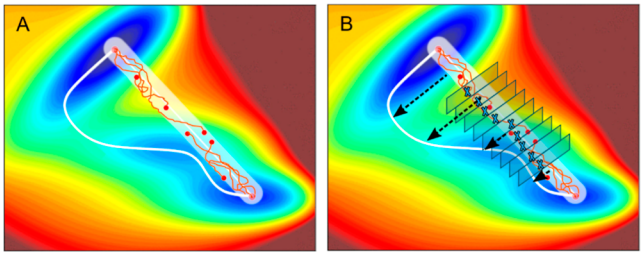


Figure 5. Illustration of ABPO. (A) Energy landscape in the CV space at the initial stage of path optimization. Multiple trajectories are launched from the two end-state energy wells (blue) and sample freely along an arbitrary initial path (red line) enhanced by V_b and within a tube centered on the path (white transparent rectangle) by a tube potential. (B) Trajectory visits to hyperplanes (small gray rectangles) perpendicular to the path tangent are counted. After sufficient sampling, the mean position in each hyperplane from counts over all replicate trajectories (blue X's) is determined, and the path and tube center are updated to these new values in CV space. The process is repeated to move the path incrementally (dashed arrows) until convergence to the optimal one (white curve).

Formulation of the ABPO path follows that of the finite-temperature SM[307] (*cf.*, Section 7.3.2). The path is specified by CVs, the definition of which is key for the computation of the PMF. Sampling of the path is counted in terms of hits to hyperplanes orthogonal to the tangent at each path index point, and statistics over multiple trajectories in a time period are used to update the ABP. The path is evolved by computing the mean position of trajectory hits in the hyperplanes and updating the path variables to coincide with those of the mean (Figure 5B). A redistribution of the updated path index points is needed for smoothing and respacing using a mollifier.[305] The optimum path is reached when the distance between the last and penultimate curves falls below a specified threshold.

The PMF $A(\lambda, t)$ along the path parametrized with λ (within an additive constant) is computed from the histograms obtained from exhaustive sampling of the optimized path over time t,

$$A(\lambda, t) = -k_B T \frac{1}{1-b} \ln\left[\frac{h(\lambda, t)}{\max[h(\lambda, t)]}\right] \tag{29}$$

As a directed approach, ABPO readily affords an atomistic description of a transition process in a reasonable simulation time depending on the choice of the selected CVs. Further, ABPO samples in a tube region surrounding the path and thereby generates a range of conformations orthogonal to the path that would not be obtained with path-restricted methods. The algorithm also provides a convenient way to assess the choice of CVs as well as the convergence of the path by following the time-course of individual CVs as a function of λ, so-called CV plots.[308,309] ABPO has the potential limitation of insufficient sampling in regions of high free energy, whereas path-restrained methods by nature ensure sampling all parts of the defined path.

7.5. Reaction Path Optimization with Holonomic Constraints. When studying protein conformational changes, a chain of intermediate replicas of the system resolve the transition between the initial and final states. To find the most probable pathway, an objective function such as the total energy or free energy of replicas is defined and minimized.[282,307] Success of reaction path optimization depends on auxiliary schemes to ensure proper distribution of replicas for capturing kinetic bottlenecks.[310] In general, it is desirable to maintain equal distances between neighboring replicas while the distance is free to change since the actual reaction path is not known *a priori*. A folded-back path should also be avoided as replicas are placed to take forward steps in crossing kinetic barriers rather than going back and forth in a basin. In this regard, the angles between three consecutive replicas are often restrained[310] to prevent drastic changes in the tangent vectors along the path that are represented by the position vector differences of replicas. A key challenge of reaction path optimization is the auxiliary scheme of managing path quality interfering with the optimization of the objective function. Keeping equal distance between replicas, for example, tends to conflict with the forces along the path in energy minimization. Although the tangential component of the force can be removed,[311] the non-conservative projected force makes the application of fast-converging gradient-based optimization methods difficult.[312] The robustness and efficiency in capturing low-energy kinetic barriers are thus limited, especially with a

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

large number of degrees of freedom and a rugged potential energy surface (PES).

The RCONS module in CHARMM overcomes this by treating equal distance between replicas as holonomic constraints.[313] Built on top of the REPLICA module, the reaction path optimization with RCONS is entirely gradient-based, readily allowing quasi-Newtonian methods and other optimization schemes assuming conservative forces. With Lagrange multipliers in constraint optimization, *ad hoc* numerical procedures such as rearranging atomic positions or force projections are not needed.[313] Furthermore, the distance between replicas can be defined by using a non-commutative RMS best-fit procedure[312] that is particularly useful for modeling transitions of macromolecules. Convergence of reaction path optimization provides a way to analyze if a sufficient number of replicas are used by testing whether the accumulated work along the optimized path agrees with the potential energy difference.[313] Since the tangent vectors in this work-energy analysis are based on positional differences between replicas, the energy or free energy difference along a path can be decomposed into contributions from different atoms to deduce the kinetic bottleneck.[314] It was also found that the straightness over replicas can be formulated as a kinetic energy potential and a temperature scale can be used to characterize the restraints regulating curvatures along the path.[313]

In principle, any potential energy function can be used to describe the energetics of replicas, and using RCONS with MSCALE provides a versatile framework for the general applications of reaction path optimization. Each replica along the path is treated as a subsystem for using a CHARMM potential energy function or in programs supported by MSCALE such as those providing a QM or QM/MM PES. For complex reactions involving conformational changes, implicit solvent model can be used to obtain an initial MEP followed by explicit-solvent MD simulations to obtain MFEP.[315,316] RCONS can also be used to constrain the sampling of MD simulation over perpendicular directions to compute the PMF along a path. In this case, the chain defined by replicas is used as a 1-dimensional order parameter for the PMF calculation.[315,316] Coupled with trajectory analysis, MD simulations constrained on the hyperplanes along a reaction path provide information about mechanistic details of a transition pathway. For example, the VIBRAN facility in CHARMM[317] can be used to compute the scale-free mechanical coupling network in proteins and nucleic acids.[318−321]

7.6. Boxed MD (BXD). BXD[322,323] is a simple technique to estimate rates and PMF $G(\rho)$ along a CV ρ in a single MD simulation. BXD falls within a class of sampling methods such as milestoning[290,324] where molecular configuration space is divided into a set of boundaries (or hypersurfaces). ρ is kept within a perfectly reflecting "box" for a time interval sufficiently long to reach convergence. This is done by reversing the velocity of the particles involved in the definition of ρ. After a given number of collisions with the boundaries, ρ is allowed to increase or decrease so that a neighboring box can be sampled. From the number of collisions with the boundaries, $G(\rho)$ over the whole range of ρ can be reconstructed, as well as the absolute rate of entering or exiting a specific "box" (Figure 6). Velocity inversion is carried out at each of the boundaries. Assuming that at a certain time the trajectory is in box m, i.e., $\rho_{m-1} < \rho(\mathbf{r}) < \rho_m$, the transition rate from box m to box $m+1$ is: $k_{m,m+1} = h_{m,m+1}/t_m$, where t_m is the time the trajectory spends in box m, and $h_{m,m+1}$ is the number of hits (i.e., velocity inversions) at the boundary between ρ_m and ρ_{m+1}. After the forward and reverse transition rates are determined, the equilibrium constants between the neighboring boxes m and $m+1$ is

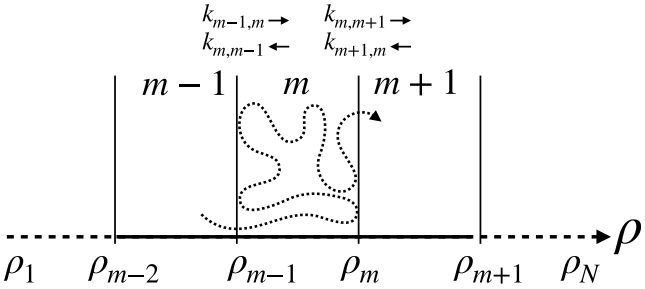


Figure 6. In BXD, the range of values assumed by the CV ρ is partitioned in boxes separated by reflective boundaries.

$$K_{m,m+1} = \frac{k_{m,m+1}}{k_{m+1,m}} = e^{-\beta \Delta G_{m,m+1}} \tag{30}$$

The free energy G_m can be determined by setting e.g., $G_1 = 0$, and the probability of finding ρ in box m is

$$p_m = \frac{\exp(-\beta G_m)}{\sum_n \exp(-\beta G_n)} \tag{31}$$

which can be multiplied by the normalized probability $P_m(\rho)$ estimated from the histograms within boxes to obtain the probability distribution function $P(\rho) = p_m P_m(\rho)$, or equivalently, $G(\rho) = -k_B T \ln P(\rho)$. In practice, the user sets the position of the boundaries and the number of times the trajectory hits a boundary before it is let into the adjacent one. Both affect the convergence, which can be assessed by performing a single simulation spanning multiple times in both directions over the range of $\rho(\mathbf{r})$ (from the lowest value to the largest, and *vice versa*). BXD is generalizable to multidimensional CVs using a general velocity-reflection procedure that conserves energy.[325]

7.7. Extended Adaptive Biasing Force (eABF) Method. Adaptive Biasing Force (ABF) is based on estimating the average force acting along a chosen CV, ξ, in order to construct and apply a biasing potential $f_m(\xi)$ that augments fluctuations of targeted dynamics.[326−328] The classical ABF method is based on TI[329] of the average force estimates which are computed in bins along the CV,

$$\begin{aligned} \Delta A &= -\int \langle F_\xi(k) \rangle_\xi \, \mathrm{d}\xi \\ &= \int \left(\left\langle \frac{\partial V(x)}{\partial \xi} \right\rangle_\xi - \frac{1}{\beta} \left\langle \frac{\partial \ln |J|}{\partial \xi} \right\rangle_\xi \right) \mathrm{d}\xi \end{aligned} \tag{32}$$

where ΔA is the free energy difference, $\langle F_\xi(k) \rangle_\xi$ is the average force along ξ in the k-th bin, $V(x)$ is the potential energy function, and $|J|$ is the determinant of the Jacobian.[327,330]

The biasing force $-\langle F_\xi(k) \rangle_\xi$ effectively flattens curvatures in the potential energy surface encountered along ξ, allowing for extensive sampling of transitions along ξ. The biasing potential is adaptive because it is updated by the current estimate of the average force along ξ until convergence.[331]

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Estimating $\langle F_\xi(k)\rangle_\xi$ brings about complications that hinder the utility of ABF.[330] For example, calculating $\partial \ln|J|/\partial\xi$ in Eq. 32 can be challenging.[332] The extended ABF method (eABF) was developed to overcome these limitations by introducing an extended potential energy function

$$V_m(x, \lambda) = V_0(x) + \frac{k_\lambda}{2}(\lambda - \xi)^2 + f_m(\lambda) \tag{33}$$

where $V_m(x, \lambda)$ is the extended potential energy, λ is a virtual particle, and k_λ is the associated spring constant.[333−336] The key distinction of eABF from ABF (Eq. 33) is the extension of the system via the λ particle; force estimates are now calculated via Hooke's law and the biasing potential is applied to λ, which augments transitions in ξ via the harmonic coupling. Since the force estimates come from the harmonic restraint between λ and ξ, the recovered PMF (along λ) may deviate from that of the physical system (along ξ) depending on the coupling strength. Several estimators have been developed to recover PMF.[331,336]

As an illustration, a simulation of gas-phase deca-alanine was performed where ξ was defined as the end-to-end distance between the terminal C_α atoms. In 500 ns, a number of transitions between the helical state and extended states are realized, with an accompanying PMF along the distance consistent with previous studies (Figure 7).[331,337]

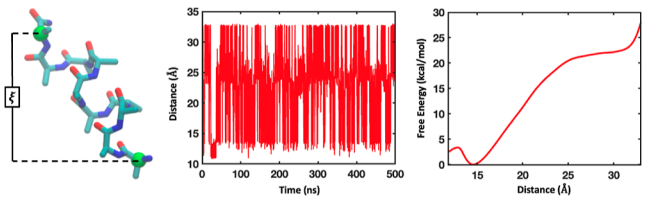


Figure 7. End-to-end distance (ξ) over time and the PMF of deca-alanine obtained using eABF.

8. ADVANCED ENERGY FUNCTIONS, COARSE GRAINING, AND IMPLICIT MODELS

8.1. Multipolar Electrostatics. Anisotropic charge distributions can be conveniently represented as a superposition of atom-centered multipoles.[338] Halogen modifications are a noteworthy example which lead to a σ-hole on the halogen atom. Such features can be represented by using multipole expansions, often up to quadrupoles.[339−343] Multipole-based electrostatics requires introducing local axes to define the orientation of higher-order multipole moments relative to the molecular geometry.

Multipolar interactions have been considered early on in molecular recognition.[344] Compared to the spherically symmetric field around a single point charge, atomic multipoles can better capture anisotropic interactions. An example is carbon monoxide that cannot be modeled well with only atom-centered point charges located at nuclear positions of the two atoms because the total charge ($Q = 0$) and the total molecular dipole $\mu = 0.048\ ea_0$ ($e = 1.6 \times 10^{-19}$ C, the charge of an electron, and $1a_0 = 0.53$ Å, the Bohr atomic length) lead to two opposite partial charges that are small in magnitude. In order to describe its substantial quadrupole moment[345−347] ($\Theta = -1.58\ ea_0^2$) either a third interaction site halfway between the two atoms is included[348] or the two atoms are described by a distributed multipole expansion.[338,349−351]

The electrostatic potential (ESP) around a molecule can be represented in general as an expansion in multipole moments where the zeroth order contribution arises from atom-centered point charges. Capturing strongly anisotropic and/or directional features, e.g., lone pairs, hydrogen bonding, π-electron density or σ-holes[352−354] requires a description beyond a single partial charge at each nuclear position. The ESP $\Phi(\mathbf{r})$ is related to the electron charge density $\rho(\mathbf{r})$ through[355]

$$4\pi\varepsilon_0\Phi(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')\,\mathrm{d}\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{Q_{lm}}{r^{l+1}}\sqrt{\frac{4\pi}{2l+1}}\,Y_{lm}(\theta, \phi) \tag{34}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are spatial variables and $1/|\mathbf{r} - \mathbf{r}'|$ was expanded in powers of $r'/r < 1$ to represent the ESP as a sum over spherical harmonics $Y_{lm}(\theta, \phi)$ from which the spherical multipole moment Q_{lm} is defined as

$$Q_{lm} = \int \mathrm{d}\mathbf{r}'\rho(\mathbf{r}')(r')^l\sqrt{\frac{4\pi}{2l+1}}\,Y_{lm}^*(\theta', \phi') \tag{35}$$

The above can be integrated to yield a compact atom-centered representation of the ESP around a molecule and are used together with the MTPL module of CHARMM.

Alternatively, multipoles of a given order can be represented by fixed charge arrangements, as is done in the distributed charge model (DCM).[356,357] It replaces the evaluation of multipole−multipole interactions with the same number of charge−charge terms at the expense of introducing additional charge sites (Figure 8A). The magnitude q_i and position of the

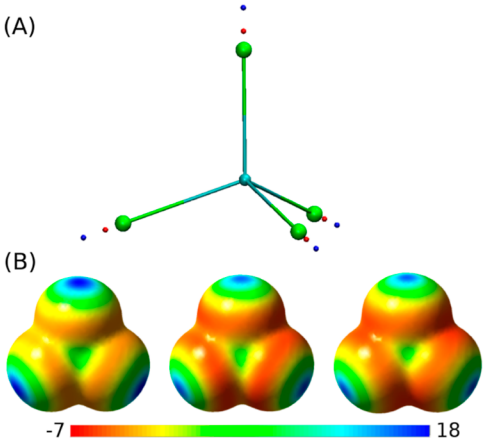


Figure 8. (A) 9-charge symmetry-constrained MDCM for CCl_4. Red and blue points respectively correspond to negative and positive charge positions. A charge at the C-atom nuclear position is hidden. (B) DFT reference ESP (kcal/[mol·e]) mapped onto the 0.001 au molecular isodensity surface (left); fitted multipolar model truncated at quadrupole (middle); 9-charge MDCM model (right, root-mean-square error (RMSE) of 0.29 kcal/mol over the grid used for fitting).

DCM charges are defined with respect to a reference atom. During fitting it may be useful to constrain the maximum displacement of the DCM charges. Reducing the number of interaction sites can be accomplished using differential evolution to optimize charge positions and magnitudes, and find arrangements that achieve a desired accuracy using a minimal number of charges (MDCM).[357] Figure 8B center and right show the ESP of a multipole representation and the

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

corresponding 9-charge MDCM model for CCl_4. More recently, the positions of the MDCM charges were explicitly coupled to the molecular geometry which leads to flexible MDCM (f-MDCM),[358] available through the DCM module in CHARMM. It can effectively capture intramolecular polarization, and meaningful atomistic simulations can be carried out for condensed-phase systems. The latest development is kernel-based MDCM (kMDCM) that uses intramolecular separations as the features in a Gaussian-kernel to describe the charge displacements depending on molecular geometry.[359]

8.2. Machine-Learning-Based Energy Functions. Over the past few years, machine learning (ML)-based approaches have flourished for constructing PESs for molecular simulations.[360−363] Typical approaches include permutationally invariant polynomials,[364,365] neural networks (NNs),[54,366,367] or kernel-based methods.[368−371] The resulting PESs have been used for gas- and condensed-phase simulations to compute observables including spectroscopic properties and reaction rates.

PES representations based on reproducing kernel Hilbert space (RKHS) have long been used for small molecules.[370,372] The TRIAKERN module in CHARMM provides the functionality to use such representations in MD simulations. The kernel coefficients $\boldsymbol{\alpha}$ required for the RKHS-based PES are determined through a versatile external utility.[371] Typical applications include reactive atom plus diatom collision systems,[373,374] but the method has also been extended to spectroscopic investigations of larger molecules.[375]

The MLpot module[376] in pyCHARMM[9] allows running of mixed ML/MM simulations. It follows more established hybrid QM/MM strategies in which a usually smaller part of the system is treated with a QM method whereas the larger remainder is represented using an empirical energy function.[377−380] In ML/MM, a ML representation, for example PhysNet,[54] is combined via mechanical embedding with an empirical FF such as CGenFF available in CHARMM.[3,17,379]

The NN-PES computes the total ML energy and forces together with electrostatic interactions between the predicted fluctuating point charges of the ML-atoms and the static atomic charges of the empirical MM atoms. CGenFF[17] handles energies and forces for the remaining MM atoms and van der Waals interactions between MM and ML atoms. Therefore, a set of van der Waals parameters must be assigned to the ML atoms. In PhysNet, charges of the ML atoms fluctuate depending on solute structure (intramolecular charge redistribution). This is akin to the fluctuating MDCM approach[358] where geometry-dependent point charges reproduce the molecular electrostatic potential and describe intramolecular polarization. The advantage of mechanical embedding is the direct application of ML-based models of atomic systems trained in the gas-phase for condensed-phase simulations without additional training. Environment-dependent electrostatics can be included at the training stage by including solvent-surrounded solutes in the training set.

The input file to run pyCHARMM (Section 3.1)[9] together with MLPot is a Python script. The MLPot module initializes an external model potential and evaluates potential energy and forces for the subset of ML atoms together with the CHARMM FF energy. By adapting the MLPot module in the source code, it is possible to link different model potentials such as ANI[381] or SchNet.[366] If the ML-based PES does not predict atomic charges, the electrostatic contribution between assigned static point charges of the ML and MM atoms are computed by the empirical energy function.

For chemical reactions that was long a domain of *ab initio* MD simulation, ML-based energy functions now provide means to run statistically significant numbers of trajectories ($\sim 10^3$ or more) which was previously not possible.[382,383] More recent examples include malonaldehyde in the gas phase,[384,385] double proton transfer in hydrated formic acid dimer,[386] or for atmospherically relevant reactions using permutationally invariant polynomials and NN-based energy functions.[382,387−389]

8.3. Multipole and Point-Induced Dipole (MPID). CHARMM now supports advanced electrostatic interactions with multipole expansion up to hexadecapole via the developmental MPOLe module. Different real-space cutoffs can be applied for PME calculations, enabling the use of smaller cutoffs for higher-order multipoles. Additionally, it allows selective exclusion of specific multipole−multipole interactions, a feature utilized for the development of a water model.[390] The module also facilitates the calculation of point induced dipoles. Several Thole damping functions, including those used in the AMOEBA[391] and MPID[392] FFs are supported. Moreover, the anisotropic atomic polarizability as utilized in the MPID model has been implemented. Induced dipole moments can be calculated either via full self-consistent field (SCF) relaxation or with an extrapolation scheme that uses weighted average of dipole moments after each of several cycles,[393] where the third-order extrapolation with empirically optimized weighting coefficients (OPT3)[394] is the default recommended method. See Section 10 for further explanation about CHARMM FFs.

8.4. Polarizable Intermolecular Potential Functions (PIPFs). The point-dipole representation of the electronic response of a molecular system to an external field offers a systematic description of polarization effects within a classical framework, where several approaches can be equivalently derived.[338] One example is the PIPF[395,396] module of CHARMM.[397,398] It complements other polarizable treatments available in CHARMM such as the fluctuating charge[399,400] and Drude oscillator[401,402] models.

In PIPF, each interaction site carries a fixed point charge and an inducible point dipole whose magnitude is determined by the total electric field due to all other point charges and induced point dipoles of the system. Assuming linear response, the induced dipole moment $\boldsymbol{\mu}_i^{\mathrm{ind}}$ at center i is proportional to the total electric field ($\mathbf{E}_i^{\mathrm{tot}}$) typically with a scalar isotropic polarizability α_i,

$$\boldsymbol{\mu}_i^{\mathrm{ind}} = \alpha_i \mathbf{E}_i^{\mathrm{tot}} = \alpha_i[\mathbf{E}_i^0 + \sum_{i \neq j} \mathbf{T}_{(2)}^{ij} \boldsymbol{\mu}_j^{\mathrm{ind}}] \tag{36}$$

The second part of Eq. 36 includes two contributions, the permanent electric field ($\mathbf{E}_i^0$) due to fixed-point charges and the induced electric field due to induced dipole moments $\boldsymbol{\mu}_j^{\mathrm{ind}}$ at other sites. $\mathbf{T}_{(n)}^{ij}$ is the rank-n polarization tensor.[338,403]

Thole's interaction dipole (TID) model[404] is employed in the PIPF model. Although an isotropic polarizability is used (a reasonable approximation for a non-interacting atom), the overall molecular polarizability is anisotropic. It has been shown that with the use of only a single parameter for each atom, the computed molecular polarizabilities using the TID model agree well with experimental data for molecules with a wide range of functional groups.[404,405]

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Due to the interdependence of $\boldsymbol{\mu}_i^{\text{ind}}$ in Eq. 36, a self-consistent iterative procedure is used in simulations to find them within a given threshold.[406] The PIPF module in CHARMM includes three complementary approaches. The first is solving Eq. 36 by direct matrix inversion.[406,407] Despite its $O(N^3)$ scaling behavior, exact results provided by direct matrix inversion are important for validating the convergence threshold for the faster iterative approach with $O(N^2)$ scaling. Direct matrix inversion is also needed for handling the intramolecular polarization part in the coupled polarization-matrix inversion and iteration (CPII) method explained below.

The next is to propagate induced dipoles dynamically via an extended Lagrangian.[408] A Verlet integrator has been implemented to couple the fictitious dipole degrees of freedom to a low-temperature bath using the Nosé−Hoover thermostat.[409−412] The low temperature dynamics makes the dipole fluctuations close to the true converged results. The extended Lagrangian method accelerates the self-consistent iteration scheme by nearly 2-fold.

The CPII approach involves a preconditioning algorithm. At convergence, the total polarization energy is

$$E^{\text{pol}} = -\frac{1}{2}\sum_i \boldsymbol{\mu}_i^{\text{ind}} \mathbf{E}_i^0 \tag{37}$$

The nuclear gradients can be obtained by differentiating Eq. 37 with respect to atomic Cartesian coordinates.[403,408] On the other hand, when direct dipole dynamics is used, evaluating energy derivatives invokes additional terms containing $\mathbf{T}_{(3)}$[408] since induced dipoles are not at the variational minimum; those third order terms can be obtained in a compact form in CHARMM.[413]

To evaluate the permanent electric field, standard nonbonded list in CHARMM is used, where intramolecular atom pairs up to 1−3 bonded pairs (connected via a single atom) are excluded and 1−4 bonded pairs (connected via two atoms) are included. To alleviate spurious polarization interactions at short-range, Thole's second damping function in the TID model[404] is used by default to determine $\mathbf{E}_i^0$ and $\mathbf{T}_{(2)}^{ij}$.

Although excluding intramolecular polarization is practical for converging induced dipoles in MD simulations, works based on the TID model suggest that both intramolecular and intermolecular polarization should be included and damped in the same way to obtain consistent molecular polarizability. Unfortunately, including intramolecular polarization between bonded pairs can introduce numerical instability in the convergence of dipoles. While it can be avoided by matrix inversion, it is computationally too expensive for the entire system. The CPII method addresses this[398] where iterative convergence of induced dipoles under intermolecular polarization is preconditioned using atom-distributed molecular polarizability tensor obtained from the TID model via matrix inversion of individual molecules. Consequently, Eq. 36 is modified to

$$\boldsymbol{\mu}_K^{\text{ind}} = \mathbf{A}_K[\mathbf{E}_K^0 + \sum_{L\neq K}^{M} \mathbf{T}_{(2)}^{KL}\boldsymbol{\mu}_L^{\text{ind}}] \quad (K = 1, ..., M) \tag{38}$$

where M is the number of molecules, K and L are indices for the corresponding atomic quantities grouped by molecules, and $\mathbf{A}_K = [\boldsymbol{\alpha}_K{}^{-1} - \mathbf{T}_{(2)}^{KK}]^{-1}$ is the atom-distributed molecular polarizability tensor[404,414] which is in units of Å^3 (see Eqs. 14 and 17 of ref 398). The CPII preconditioning algorithm has been implemented in CHARMM and it accelerates dipole convergence in liquid simulations of amide and polypeptide systems.

In addition to gradients, second derivatives of PIPFs can be calculated in CHARMM with the point dipole formalism.[403] With analytical Hessian available, PIPFs can be used in conjunction with the VIBRan module in CHARMM for vibrational FF analysis.

The PIPF model has been employed in MD simulations to examine polarization effects in a series of organic liquids including alkanes, alcohols, and amides.[395−397] The results obtained with the classical point-dipole model were found to be in good agreement with those from combined QM/MM simulations in which polarization effects are described quantum mechanically.

Recently, the PIPF model has been employed in the doubly polarized QM/MM (dp-QM/MM) method to enhance the accuracy of SE-QM/MM (SE: semiempirical) free energy simulations.[415] A well-known limitation of SE-QM methods is their tendency to underestimate molecular polarizability compared with experiments and AI/DFT-QM (AI: *ab initio*) benchmarks, leading to significant errors in free energy profile determined at SE-QM/MM levels. The dp-QM/MM method addresses this by improving the response properties of SE-QM/MM methods through high-level molecular polarizability fitting. Specifically, additional induced point dipoles are introduced on the QM atoms through a set of corrective polarizabilities ("chaperone polarizabilities"), whose magnitudes are determined from ML to reproduce the condensed-phase AI-DFT molecular polarizability along the MEP. These chaperone polarizabilities are then used in PIPF calculations in conjunction with QM/MM to compensate for the polarization energy underestimate in conventional SE-QM/MM simulations. Applied to the Menshutkin reaction in water, the dp-QM/MM method brought the computed and experimental free energy results into closer agreement.[415]

8.5. Long-Range Lennard-Jones Interactions. While the PME method[35,176] for evaluating long-range electrostatic interactions was added to CHARMM in 1995,[416] implementing the long-range Lennard-Jones (LJ) interaction lagged considerably. Long-range effects of dispersion are important for accurate calculation of free energies and interfacial properties of liquids and surfactants.[417−419] Ignoring them leads to inconsistencies in the surface tension of lipid bilayers and monolayers.[420] To address long-range LJ interactions, a lattice-based method termed LJ-PME[421] was implemented into CHARMM.[422] The name LJ-PME is arguably a misnomer, in that only the C6 (r^{-6}) dispersion is calculated with an Ewald summation. The C12 (r^{-12}) term continues to be truncated with a standard switching function.[4] Dispersion-PME has been used interchangeably and is perhaps a better name.

Key to the efficiency of electrostatic PME is the simple multiplicatively separable functional form. While the analogous PME method for dispersion has been available for decades,[176,423] its adoption has likely been hindered by the fact that many FFs, including CHARMM, use the Lorentz rule

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{39}$$

to combine the LJ geometric parameters σ_i and σ_j for distinct atom types i and j. Since this does not yield a multiplicatively separable form, binomial expansion can be used:

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

$$(\sigma_i + \sigma_j)^6 = \sum_{k=0}^{6} \binom{6}{k} \sigma_i^k \sigma_j^{6-k} \tag{40}$$

However, this approach requires several PME evaluations in addition to those used in electrostatic PME. An elegant solution[421] is to instead assume geometric mean combination rule

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j} \tag{41}$$

and apply PME to the dispersion part of the LJ potential, requiring similar effort to that employed in the electrostatic term. This allows the C6 term to be calculated with a single PME calculation involving a simple multiplicative form, $C6_{ij} = \sqrt{C6_i C6_j}$.

To correct for this approximation, the geometric-mean term is analytically subtracted (similar to how 1−2 and 1−3 exclusions are handled in electrostatics) for all pairs within a cutoff distance and substituted with the correct form, which may possibly include NBFIX pair-specific corrections (Section 10). The net effect is that the LJ potential is exact up to the chosen cutoff, beyond which the geometric-mean functional form is used. For an ∼8-Å cutoff, the geometric- and arithmetic-mean potentials coincide very closely and the overall approximation is excellent. To account for small but abrupt changes in energy as two schemes handoff at the cutoff, the original formulation in CHARMM applied a shift term to ensure continuity in energy. However this approach does not guarantee continuity in the first derivative. For this reason, a sigmoidal switch function has been implemented to seamlessly transition between the two regimes.[422]

The C6 terms are generally computed using the van der Waals parameters in CHARMM. To achieve greater flexibility, these values may be input directly regardless of atom types or short-range C6 values. In this way, specific interactions can be "fixed" to simplify parameter fitting and optimization by treating short-range and long-range dispersion terms independently.

The LJ-PME algorithm provided the impetus to revisit the use of cutoffs in the CHARMM lipid FF (Section 10.4). It has long been known that the high anisotropy present in such systems renders the standard isotropic corrections inappropriate. Because PME makes no assumption about the isotropy of the system, it is well suited to general systems including lipids. Previous iterations of the CHARMM lipid FF were parametrized with a given cutoff, with the cutoff errors implicitly absorbed into the parametrization. While successful, this strategy makes the resulting FF sensitive to the cutoff used at runtime; a choice other than that used for parametrization can yield erroneous results.

Long-range LJ terms can also be evaluated in CHARMM using the Isotropic Periodic Sum (IPS)[424] and extended IPS methods.[417,418] However, LJ-PME provides higher efficiency and transferability of FF among different simulation programs.

8.6. FACTS Implicit Solvent. The Fast Analytical Continuum Treatment of Solvation (FACTS) model is an efficient GB-based implicit solvent method for calculating the solvation free energy of proteins, protein complexes, and protein−ligand interactions.[132] FACTS is based on the analytical evaluation of the volume and spatial symmetry of the solvent that is displaced from a solute atom by nearby atoms. For each solute atom, these two measures of solvent displacement are combined into an empirical sigmoidal equation for the calculation of the atomic (or self) electrostatic solvation energy and the solvent accessible surface area (SASA). The former is used to calculate the Born radii in the GB equation. The SASA is used to evaluate the non-polar contribution to solvation.

FACTS is fully analytical and because of its speed, it is useful for MD simulations. It has two main advantages over other implementations of the GB model. First, FACTS does not use the so-called Coulomb field approximation where the electric displacement field of a solute atom is calculated by assuming that the solute−solvent dielectric boundary is spherical with the atom at the center of the sphere. This assumption breaks down particularly for solute molecules with substantial aspherical volume and/or charges located close to the solute−solvent boundary. Second, FACTS does not require setting the dielectric discontinuity surface. Importantly, the CHARMM energy calculation with the FACTS model is only about four times slower than the vacuum energy, and FACTS scales linearly with system size (see Figure 11 in ref 132).

FACTS is versatile as parameters for new or unknown atom types (e.g., non-proteinaceous atoms in organic compounds) can be generated automatically by interpolation from the existing FACTS parameters by the FACTS keyword TAVW. The effect of salt (ionic strength) is treated by the linearized Debye−Hückel approximation.[425] The FACTS energy and force terms have been parallelized for multiple CPUs which provides substantial speed-up particularly for large systems. Furthermore, FACTS is compatible with the IMAGE module for periodic boundary condition (PBC) and the BLOCK module for energy decomposition. The INTE command of CHARMM can be used to evaluate the FACTS energy between two groups of solute atoms, e.g., a protein and a small-molecule ligand.

Since the original publication in 2008, FACTS has been employed in many simulation studies of (small) protein folding, (poly)peptide amyloid aggregation, and binding of ligands to proteins. One interesting example is the MD study of the interactions of the toxic Alzheimer's $A\beta_{1-42}$ peptide with carnosine, the endogenous brain dipeptide β-Ala-His, which revealed salt bridges with charged side chains, and van der Waals contacts with residues in and around the ^{17}LVFFA^{21} central hydrophobic cluster of $A\beta_{1-42}$.[426] In 2014, the FACTS model was extended to lipid bilayer (membrane) environment by using a position-dependent dielectric constant and an empirical surface tension parameter. It was shown to reproduce the self-energy and pairwise interaction energies in solution calculated by the finite-difference Poisson method.[427] However, the FACTS model for the membrane has not been implemented into the official version of CHARMM yet.

The last sentence of the original FACTS paper mentioned potential applications beyond MD: "*The accuracy and efficiency of FACTS suggest that it could also be used for protein structure prediction and docking*".[132] In fact, FACTS has been employed in several docking programs. A recent example is the FASTDock pipeline for the efficient scoring of poses in ortho- and allosteric pockets generated by MD.[127] Another example is the docking protocol called Attractive Cavities which uses energy minimization and a smoothed potential energy for guiding small molecules into protein cavities. In a successive refinement step, the binding energy is calculated as the sum based on the CHARMM FF and the FACTS model.[428,429] Concerning structure prediction, a Python tool allows for the evaluation of the FACTS total energy and its

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

contribution as descriptors for ML models of protein−protein interactions.[430]

8.7. Implicit Modeling of Membranes. *8.7.1. Membrane Pores.* The IMM1 implicit membrane model,[431,432] an extension of the EEF1 effective energy function for soluble proteins,[433] has been adapted to account for aqueous pores.[434,435] IMM1 uses two sets of solvation parameters, one for water and one for the non-polar membrane interior. A continuous switching function f describes the transition from one environment to the other. Modeling of pores is accomplished by using a switching function F dependent on the vertical (z) and radial coordinate (the distance r from the z-axis). Denoting the thickness of the non-polar part of the membrane as T and the pore radius as R (Figure 9), the

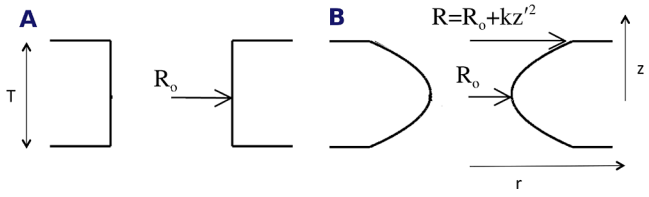


Figure 9. Illustration of pore shapes in IMM1. (A) Constant radius. (B) Radius depending on the z-coordinate.

switching functions can be expressed using dimensionless variables:

$$F(z', r') = f(z') + b(r') - f(z')b(r'),$$
$$f(z') = \frac{z'^n}{1 + z'^n},\ b(r') = 1 - f(r'),\ z' = |z|\frac{2}{T},\ r' = \frac{r}{R} \quad (42)$$

Different pore shapes can be modeled by making R dependent on z, e.g., $R = R_o + kz'^2$ (Figure 9B). Because the Gouy−Chapman formulas are invalid for pores in anionic membranes, an alternative approach is based on numerical solution of the Poisson−Boltzmann equation.[436] It has been used to investigate the pore forming activity of antimicrobial peptides.[435,437−442] More recently, this model was used for initial evaluation of putative structures of β barrel membrane pores formed by fibril-forming peptides and proteins, such as amyloid β, IAPP, and α-synuclein.[443−446] A similar energy function has been implemented in the Rosetta protein design package.[447]

8.7.2. Curved Membranes. The IMM1 implicit membrane model has been extended to spherical and cylindrical membranes (vesicles and tubes) by changing the definition of the relative depth z' from $|z|/(T/2)$ to $|r - R|/(T/2)$ where R is the radius of the vesicle or tube and r the radial position of an atom.[448] The model can also account for changes in lateral pressure profile as the membrane bends.[449] It has been used to study ESCRT-III snf7,[450] the mechanism of negative curvature generation by IBAR domains,[451] and the interaction of caveolin oligomers with membranes.[452]

8.7.3. Mean-Field Modeling of Deformable Membrane Bilayers via HDGB. The Heterogeneous Dielectric Generalized Born (HDGB) model[453] is an extension of the GBMV implicit solvent model[160] to capture the interaction of biomolecules with biological membranes. The most straightforward approach for implicitly modeling membrane−water interfaces is via a two-dielectric system where the membrane is modeled as a low-dielectric slab (ϵ = 1−2) embedded in a high-dielectric region (ϵ = 80). This idea was implemented in earlier implicit membrane models.[229,431,458] The HDGB model refines it by introducing a continuously varying dielectric profile across the membrane−water interface to better describe the actual dielectric profile of membrane bilayers.[459] The variable dielectric profile is then used in a modified GB equation

$$\Delta G_{solv}^{GB} = -166\sum_{i\neq j}^{N}\left[1 - \frac{2}{\epsilon_i + \epsilon_j}\right] \times \frac{q_i q_j}{(r_{ij}^2 + \alpha(\epsilon_i)\alpha(\epsilon_j)e^{-r_{ij}^2/[F\alpha(\epsilon_i)\alpha(\epsilon_j)]})^{1/2}} \quad (43)$$

where $166 = (8\pi\epsilon_0)^{-1}$ (ϵ_0: permittivity of vacuum) is a factor arising from units used in CHARMM, N is the number of atoms, ϵ_i is the variable dielectric profile at the position of atom i with charge q_i, typically along the z-direction coinciding with the membrane normal,[453,460] r_{ij} is the distance between atoms i and j, $\alpha(\epsilon_i)$ are Born radii dependent on ϵ_i, and F = 8 is a dimensionless parameter. In addition to a variable dielectric profile, the HDGB model introduces a SASA-dependent non-polar contribution:

$$\Delta G_{cavity} = \sum_{i=1}^{N} SASA_i \gamma S(z) \quad (44)$$

where $SASA_i$ is the SASA of atom i, and γ is the surface tension reflecting the strength of the non-polar term, and $S(z)$ is an optimizable profile along z implemented as a spline-interpolated function. In principle, HDGB can be applied to any heterogeneous dielectric environment including those with non-slab geometries. For example, spherical micelles can be modeled by varying the dielectric and non-polar profiles as a function of the radial position from the center. It is also possible to exclude select atoms from the variable dielectric and non-polar profiles, which allows modeling of membrane channels where membrane-facing atoms would be in contact with water or ions instead of the lipid bilayer.

HDGB variants available in CHARMM are summarized in Table 2. In the original HDGB,[453] the dielectric profile was

Table 2. HDGB Model Variants Implemented in CHARMM[a]

Model	Dielectric Profile	Non-Polar Profile	VDW Term	Deformable
HDGB[453]	PB calculations for multilayer dielectric	Insertion profiles of O_2 and water	No	No
HDGBv2[454]	ΔG_{insert} for amino acid analogues		No	No
HDGBv3[455]	Intramembrane interactions and ΔG_{insert} for amino acid analogues		No	No
HDGBvdW[456]	Same as HDGBv3	Reoptimized as for HDGBv3	Yes	No
DHDGB[457]	Any	Any	No	Yes

[a]PB, Poisson−Boltzmann; ΔG_{insert}, insertion free energy of amino acid analogues; VDW, van der Waals.

taken from Poisson−Boltzmann calculations for a probe sphere at different locations in a multilayer dielectric system and the non-polar profile was adjusted to match insertion free energies of model molecules. In HDGBv2,[454] both dielectric and non-polar profiles were optimized further to improve insertion free energies of amino acid analogues. In HDGBv3,[455] side-chain

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

interactions within the membrane from all-atom simulations were taken into account. HDGBvdW[456] includes implicit van der Waals interactions inside and outside of the membrane separate from the cavity non-polar term. This further improved intramembrane interactions. Furthermore, DHDGB[457] adds dynamically fluctuating membrane deformations by coupling HDGB to membrane deformation energies from elasticity theory. This allows for a more realistic modeling of charged and polar compounds near and inside the membrane bilayer via membrane deformations. The DHDGB model can be combined with any other HDGB models in Table 2.

With HDGB, it is possible to study a variety of peptide−membrane interactions. A significant advantage of an implicit membrane model is that slow bilayer reorganization kinetics can be avoided. This is especially relevant for peptide−membrane insertion where atomistic simulations may be too slow to converge. HDGB has been used successfully to study the insertion of viral fusion peptides.[461,462] Another advantage is that the width of the bilayer can be easily varied simply by scaling the dielectric and non-polar profiles. This allowed a comparison of phospholamban conformational sampling in different physiologically relevant bilayers.[454] It also led to a method for estimating the optimal membrane width based on the structure of a given integral membrane protein.[463]

HDGB was also used to estimate the membrane permeability of drug-like molecules,[464] and as a scoring function for membrane protein structures,[465] which led to a MD-based structure refinement protocol for integral membrane proteins.[466] HDGB can also be used for simply simulating the dynamics of membrane-embedded integral proteins.[467] However, for larger systems, the computational advantage of HDGB over explicit lipid simulations is not as significant.[468]

8.8. Transferable Coarse-Graining via PRIMO. The Protein Intermediate Resolution MOdel (PRIMO) is a CG model for proteins and nucleic acids with resolution intermediate between atomistic and residue levels.[469] In PRIMO, protein backbones are represented with three particles: C_α, N, and CO (at the midpoint of the carbonyl group). Non-glycine side chains are represented with one to five beads depending on the size of the side chain (e.g., Figure 10A). Nucleic acids are represented with a similar level of coarse-graining.[469]

The CG sites in PRIMO were chosen to allow an analytical reconstruction of atomistic detail with minimal loss of accuracy by applying known standard bond geometries.[469] On average, all-atom reconstructions from PRIMO deviate by only 0.1 Å from the original all-atom models.[469] For comparison, at the time PRIMO was developed, all-atom reconstructions from C_α-only models deviated by 1.7 Å[469] on average and even when C_α sites were combined with a site at the side chain center, the reconstruction error remained at 0.9 Å.[469] More accurate all-atom reconstructions from residue-level CG models up to 0.5 Å are now possible with advanced ML models,[470] but PRIMO has an advantage by maintaining very close connection to atomistic models with significantly reduced number of interaction sites. The near-exact mapping between CG and atomistic levels in PRIMO can be used to compress all-atom trajectory data.[471]

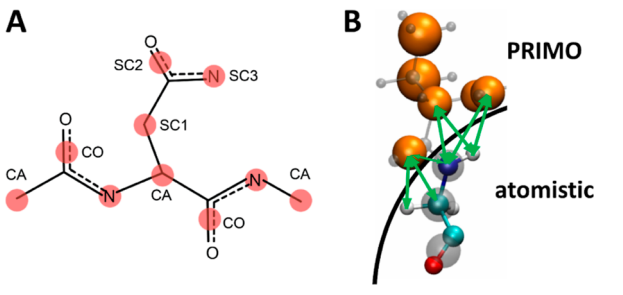


Figure 10. Illustration of the PRIMO CG model. (A) PRIMO interaction sites (red spheres) for asparagine as an example. (B) Hybrid all-atom/CG model of a protein with coupling between a PRIMO region and an atomistic region.

The PRIMO FF incorporates a combination of bonded and non-bonded terms in all-atom FF.[472] It augments standard bonded terms with spline-based functions because interactions between many of the CG sites have multiple minima that are not approximated well with single-well harmonic terms. Moreover, some bonded terms operate on virtual sites that are reconstructed on-the-fly from the CG sites, which is possible because of the computationally efficient analytical mapping from PRIMO to atomistic sites. The virtual site approach is also used for an explicit hydrogen-bonding potential using 2D spline interpolations of PMF as a function of hydrogen bond angle and distance similar to the CMAP torsion potential in CHARMM.[473] Finally, PRIMO captures solvation effects via implicit solvent terms. A GBMV-based model[160] is used to capture electrostatic contributions to the solvation free energy. It is complemented with a per-residue SASA term that adds non-polar contributions and compensates for incomplete electrostatic solvation contributions with the GB model because of less polarized PRIMO interaction sites. By replacing GBMV with the HDGB implicit membrane model,[453] PRIMO can be extended to simulate protein−membrane interactions.[474] The PRIMO FF were parametrized primarily by matching energies from CHARMM's all-atom FFs, in particular CHARMM22/CMAP[475] and CHARMM36[476] with further adjustments made based on simulations of test peptides.[472]

The PRIMO FF is fully transferable to other systems.[477] It is possible to run stable MD simulations of arbitrary protein systems without restraints[472,474] and PRIMO can be used in combination with enhanced sampling techniques in peptide folding simulations,[472] or to study the insertion of peptides into membranes.[474] PRIMO is also useful for extensive conformational sampling in protein structure refinement.[478]

Because PRIMO and atomistic interaction potentials are similar and compatible with each other, it is possible to run hybrid multiscale simulations where parts of a system are represented by PRIMO whereas other parts are represented in atomistic detail. One example is the simulation of a peptide in atomistic detail surrounded by crowder molecules represented at the CG level using PRIMO,[479] where the coupling between the CG and atomistic levels involves only the non-bonded and solvation terms. It is also possible to run multiscale simulations where different parts of the same molecule are represented either atomistically or via PRIMO.[480] The coupling between the all-atom and CG levels extends to the bonded terms by maintaining dual resolution across the interface between CG and atomistic regions. PRIMO resolution is trivially obtained from the atomistic level whereas atomistic sites are reconstructed analytically from the PRIMO model (Figure 10B).[480] Using this approach, it is possible, for example, to efficiently sample dynamic regions of a given system at the CG

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

level while applying atomistic detail to maintain accuracy of more conserved structural elements.[480]

PRIMO is unique because of its close coupling to the CHARMM all-atom FF. Its advantage is a high degree of transferability compared to other CG models and a suitability for multiscale simulation approaches where a given system can be represented simultaneously at different levels of resolution. The CG nature of PRIMO improves computational efficiency over comparable all-atom simulations with the same GBMV implicit solvent model by about a factor of 10 in wall-time.[472] There are additional gains in efficiency due to the smoother energy landscape at the CG level. However, the use of the relatively expensive GBMV model limits PRIMO's overall performance, especially for larger systems where the advantage of implicit solvent over explicit solvent diminishes.[468] Combining PRIMO with other less expensive implicit solvent models in CHARMM is an option for potentially overcoming these limitations.

9. SPECIALIZED RESTRAINT METHODS

Restraint energy functions apply biases on particular degrees of freedom, and can drive the system into conformational states that may be otherwise inaccessible, thereby improving sampling around those states.

9.1. CONSHELIX Module. Most restraint potentials control reaction coordinates between atoms, such as CONS HARM, NOE, and RESD for atoms, and CONS DIHE for dihedral angles in CHARMM. By comparison, the restraint potentials in the CONSHELIX module are applied to molecular-level reaction coordinates such as helices and hairpins (Figure 11). These energy functions are especially

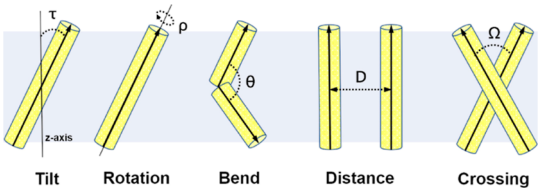


Figure 11. Helical reaction coordinates.[481,482] The tilt angle τ is relative to the z-axis; the rotation angle ρ is measured for a designated atom about the helical axis; the bend angle θ is between two helical axes; the minimum distance D and crossing angle Ω are between two neighboring helices. Corresponding restraint energy functions are handled by the CONSHELIX module in CHARMM. For a hairpin, tilt, rotation, and distance restraint potentials are also available.

useful for controlling motions of transmembrane domains within the lipid bilayer,[483,484] which are critical for, e.g., signal transduction,[485,486] transport of ions and small molecules, antimicrobial activity, and transmembrane responses.[487]

The CONSHELIX restraint potential $U_\xi(\mathbf{R})$ (Eq. 45) takes a quadratic form for non-periodic variables ($\xi = \tau$, θ, D, or Ω; defined in Figure 11) and a cosine function for periodic variables ($\xi = \rho$), where $\mathbf{R}$ represents coordinates of atoms selected to define the helix/hairpin principal axis. Denoting the force constant and the target value as k_ξ and ξ_0, respectively,

$$U_\xi(\mathbf{R}) \begin{cases} \frac{1}{2}k_\xi[\xi(\mathbf{R}) - \xi_0]^2 & \text{(non-periodic)} \\ k_\xi[1 - \cos(\xi(\mathbf{R}) - \xi_0)] & \text{(periodic)} \end{cases} \tag{45}$$

An example application is the study of the mismatch in the thickness of the hydrophobic region between the protein and the lipid bilayer,[487] which leads to changes in lipid length, tilting of transmembrane proteins, and association of multiple transmembrane proteins. To explore the resultant tilting motion of the WALP19 model helix peptide (sequence: GWW(LA)$_6$LWWA) in a DMPC bilayer, umbrella sampling was performed.[488,489] The weighted histogram analysis method (WHAM) and TI were utilized to obtain the PMF as a function of the tilt angle τ, where 'precession entropy' was proposed as the driving force for tilting. As τ increases, the accessible volume of the helix conformation also increases (precession entropy), which stabilizes the helix in a tilted orientation.

Another application is the association free energy between two helices using the distance D in a DMPC bilayer.[490] The asparagine residue at the center of the helix drives helix−helix association in a membrane by forming bifurcated hydrogen bonding. However, the interaction between the helix and the lipid promotes dissociation of the two helices. The free energy cost for lipid depletion between helices is relatively small, around 7.6 cal/[mol·Å^3], compared to that for cavity formation in water, 24−33 cal/[mol·Å^3]. Through TI, various energy terms within a residue can be decomposed, which inform residues that favor helix−helix or helix−lipid interactions.

9.2. SSNMR Module. Solid-state NMR (SSNMR) spectroscopy is used to determine membrane protein structures in a native-like membrane environment. It utilizes 2D ^{15}N−^{1}H NMR polarization inversion spin exchange at magic angle (PISEMA) spectrum experiments to obtain orientational information from dipolar coupling and chemical shift. To harness this experimental data effectively, the SSNMR module has been implemented in CHARMM.[491]

Experimental observables (O_{ssnmr}) obtained through the SSNMR spectroscopy are dipolar coupling (ν) for ^{15}N−^{1}H pair and chemical shift (σ) for ^{15}N atom:

$$\nu = \frac{\nu_0}{2}(3\cos^2\theta - 1)$$
$$\sigma = \left\langle \sum_{i=1}^{3} [\hat{n}\cdot\hat{e}_i(t)]\sigma_{ii}(t)[\hat{e}_i(t)\cdot\hat{n}] \right\rangle \tag{46}$$

Here, $\nu_0 \equiv \gamma_N\gamma_H h\mu_0/(8\pi r^3)$ is the dipolar coupling constant, where $\gamma_N = -2.71 \times 10^7/(\text{T}\cdot\text{sec})$, $\gamma_H = 2.675 \times 10^8/(\text{T}\cdot\text{sec})$, h is the Planck constant, μ_0 is the vacuum permeability, and r is the N−H bond length. θ is the angle between the N → H bond vector and the direction n̂ of the magnetic field that is assumed to be normal to the membrane. $\hat{e}_i$ (i = 1, 2, 3) are the instantaneous basis vectors for the chemical shift tensor where σ_{ii} is its diagonal element. $\hat{e}_1$ and $\hat{e}_3$ are on the plane spanned by C, H, and N atoms of the peptide bond, and $\hat{e}_2 = \hat{e}_3 \times \hat{e}_1$.[491−493]

Due to $\cos^2\theta$ in Eq. 46, for one dipolar coupling experimental value, up to four helix orientations are possible, potentially with multiple helical structural conformers in membrane ('structural ambiguity'). The simplest form of restraint energy U is a quadratic function that minimizes experimental values

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

$$U = \begin{cases} \sum_{i}^{M} k_{dc}(|\nu_i| - \nu_{exp,i})^2 & \text{(dipolar coupling)} \\ \sum_{i}^{M} k_{cs}(\sigma_i - \sigma_{exp,i})^2 & \text{(chemical shift)} \end{cases} \quad (47)$$

where M is the number of restraint potentials and the subscript 'exp' denotes experimental value. Eq. 47 can be minimized through various methods such as MD, MC, genetic algorithm optimization, and simulated annealing. A structure that satisfies experimental data can thereby be obtained. An example is modeling the structure of the fd-coat, the major pVIII coat protein of the fd filamentous bacteriophage (Figure 12). It has

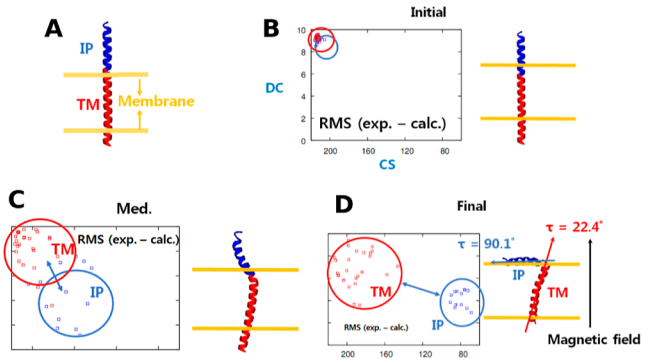


Figure 12. Structural changes of the fd-coat membrane protein through MD simulations with the SSNMR restraint potential. (A) The fd-coat system. IP (blue): In-plane helix; TM (red): Transmembrane helix. (B) Distribution of chemical shift (x-axis) and dipolar coupling (y-axis) calculated for the initial state structure. (C) Structure and calculated distribution in an intermediate stage of the simulation. Note changes in the distributions and conformations. (D) The final stage of the simulation. IP is at the lipid–water interface, while TM is positioned within the membrane.

two types of α-helices within the lipid membrane: the amphipathic in-plane (IP) helix at the water–lipid interface, and the longer transmembrane (TM) helix. Solution and solid-state NMR structures are available (PDB ID: 1FDM and 1MZT).[494] Calculations were performed in vacuum while considering a virtual lipid environment where Eq. 47 was used with experimental values as minima. Initially, the structure is perpendicular to the lipid membrane. Subsequently the IP region adheres to the lipid membrane surface, and the TM region adopts a tilted structure to match the hydrophobic length of TM and the membrane thickness (Figure 12), which satisfy constraints based on experimental data.

Other applications of the SSNMR module on transmembrane proteins include MerF (a mercuric ion transporter), M2 (transmembrane domain from influenza A virus), and Vpu (viral protein u from HIV-1).[491] In another application, the SSNMR module was combined with the ensemble dynamics (ED) technique[495−498] in an explicit membrane system to address discrepancies between semistatic and dynamic fitting models of SSNMR observables. Compared to these two fitting approaches, the main advantage of the SSNMR ED is its ability to generate an ensemble of structures (e.g., TM helix orientational distribution) that satisfy experimental observables within a reasonable physical model, without prior knowledge about the underlying distribution or motion.

9.3. Residual Dipolar Coupling (RDC) NMR Orientational Restraint. The RDC module in CHARMM leverages experimental time-averaged RDC orientational NMR restraints (Figure 13).[499] RDC informs about the orientation of each

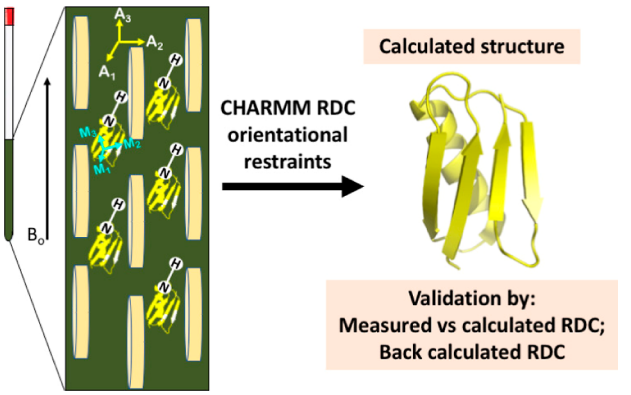


Figure 13. From RDC (D_{PQ}) measurement to structure calculation using CHARMM illustrated by using the N–H internuclear vector of protein G (PDB 1P7E).[500] The protein solution in the NMR tube is indicted in green. Oval-shaped slabs indicate the alignment media. $\mathbf{B}_0$ is the applied static magnetic field, and $\mathbf{A}_i$ and $\mathbf{M}_i$ (i = 1, 2, 3) represent the 3 principal axes for the alignment tensor and the inertia tensor of the molecule, respectively.

internuclear vector $\mathbf{r}_{PQ} = \mathbf{r}_P - \mathbf{r}_Q$ formed by a pair of NMR active nuclei P and Q in a molecule with respect to the static magnetic field $\mathbf{B}_0$. CHARMM uses decoupled RDC orientational information[500] that consists of (1) the angle ψ_i between $\mathbf{B}_0$ and the i-th principal axis of the inertia tensor of the molecule, $\mathbf{M}_i$, and (2) the angle θ_i formed between $\mathbf{r}_{PQ}$ and $\mathbf{M}_i$. The RDC between the two nuclei is

$$D_{PQ} = \frac{2}{3}\frac{D_{const}}{r_{PQ}^3}\sum_{i,j=1}^{3}\left\langle \frac{3}{2}\cos\psi_i\cos\psi_j - \frac{1}{2}\right\rangle \times \left\langle \frac{3}{2}\cos\theta_i\cos\theta_j - \frac{1}{2}\right\rangle \quad (48)$$

where $D_{const} = -S\mu_0\gamma_P\gamma_Q h/8\pi^3$, where μ_0 is the magnetic permeability of vacuum, γ_P and γ_Q are gyromagnetic ratios of the nuclei P and Q, h is the Planck constant, and S is the generalized order parameter that describes the internal motion of the internuclear vector. The angular brackets indicate time average. Eq. 48 can be expressed in terms of the alignment tensor or Saupe order matrix A

$$D_{PQ} = \frac{D_{const}}{r_{PQ}^5}\mathrm{Tr}[AO^T R_{PQ} O] \quad (49)$$

where O consists of the three principal axes (eigenvectors) of $\mathbf{M}_i$ and the 3 × 3 matrix $R_{PQ} = \mathbf{r}_{PQ} \otimes \mathbf{r}_{PQ}$. Since A is traceless, only 5 components are independent, which can be determined by using singular value decomposition (SVD) with the aid of experimental RDC,[501] $D^{exp}_{n\times 1}$ (n: number of experimental values) and $\mathbf{M}_i$[500]

$$\hat{A}_{5\times 1} = V_{5\times 5}[1/W]_{5\times 5}U^T_{5\times n}D^{exp}_{n\times 1} \quad (50)$$

where V, $1/W$, U and D are matrices arising from SVD.

CHARMM supports simultaneous use of RDCs measured between different NMR active nuclei (e.g., N–H, C_α–C, and C–N). Since RDC provides only the orientational informa-

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

tion, it alone cannot fully determine the structure of a biomolecule. Thus, RDC and other NMR restraints such as NOE (via the NOE keyword) are used together. Alternatively, one can use a fragment-based approach[502] to determine molecular structure using multiple RDCs collected in multiple alignment media. It is recommended to "reset" the RDC module before using it for structure determination.

The module requires the name of the input file(s) containing experimental RDC values, their upper and lower bounds, and the corresponding internuclear atoms. Since the CHARMM RDC module uses the harmonic potential with soft asymptotic behavior, for each RDC, one can specify the force constant for the harmonic potential and the slope for the asymptotic function (default = 1), the value for the exponential function used in the soft asymptote if it is other than 1, and the cutoff length for the harmonic function (default: 1 Å). The input files may contain RDCs collected in different alignment media, stored in the CHARMM (default), BMRB, or XPLOR format. Other command options include the maximum RDC restraints, whether the RDCs (other than for N−H) need to be scaled with respect to the N−H RDC, and whether the principal axes are calculated only with respect to the RDC atoms. For validation, users can compare experimental versus calculated RDC data and back-calculate other types of RDC (e.g., C−N and/or C_α−C) for the structure based on the input RDC used (e.g., N−H).

9.4. Torque Application. Application of a torque on selected atoms about a user specified axis can be achieved by the PULL TORQue command.[503]

10. CHARMM FORCE FIELD DEVELOPMENT

Since the publication of the second CHARMM paper in 2009,[3] significant progress has been made in the CHARMM-related FFs. While the additive CHARMM36 (C36) FF was mature at that time, additional enhancements and refinements, including creation of CGenFF, were performed. With respect to polarizable FFs, major developments in the classical Drude oscillator model were made including coverage of all the major classes of biomolecules and progress was made toward a Drude General FF (DGenFF). A detailed description of the potential energy functions for the additive and Drude FFs is presented in ref 23.

In addition, a polarizable model based on MPID (Section 8.3) onto which Drude FF-based parameters can be mapped was presented.[392] To avoid the costly SCF procedure (Section 8.3), a small mass is assigned to the Drude particles, which are then propagated as dynamic variables during simulations via a dual-thermostat extended Lagrangian algorithm, with a "cold" temperature imposed on the degrees of freedom corresponding to the induced dipole. The statistical mechanical validity of this procedure was clarified.[504] In addition to CHARMM, the additive and Drude FFs are available in OpenMM,[505] facilitating the CHARMM/OpenMM API (Section 2.1), as well as in GROMACS[506] and NAMD,[507] though the implementation of the Drude FF in GROMACS is currently limited, and NAMD currently does not include LJ-PME capabilities (Section 8.5).[421] Notable is the ability to setup and generate inputs for complex molecules for a range of programs using CHARMM-GUI[87] (Section 3.3), including for the Drude polarizable FF.[115] An advantage of the Drude FF over other polarizable FFs is the computational efficiency, i.e., it is only about 4 time slower to run compared to additive FFs.

10.1. Water, Ions, and Polar Solvents. A number of extensions of the Drude FF, including additions to the potential energy function, have been made with respect to water and ions since 2009. Extensions to the additive FF include the use of alternative LJ interactions between the CHARMM TIP3P water model and proteins[144] (Section 10.2), and revised LJ parameters for Na^+ and Ca^{2+} with lipids.[508,509] The default water model with the Drude FF is the SWM4-NDP (simple water model with 4 sites and negative Drude polarization) model,[510] though a 6-point model SWM6 with an improved condensed phase hydrogen bonding properties is available with an increased computational cost.[511] The energy function was expanded to account for polarization anisotropy to describe the dielectric constant of liquid amides more accurately.[512] Substantial work was undertaken on the ions in the Drude FF, including a range of monatomic ions[513,514] and later for molecular ions, including a number of ions uncommon in biomolecular systems.[515−517]

Developing the parametrization for all charged moieties made it possible to adopt a consistent absolute solvation scale for monatomic and molecular ions.[513,515] Notable with the optimization of Mg^{2+} was the use of an LJ repulsion between the SWM4 Drude particle and the Mg^{2+} ion.[517] This enabled steric repulsion between water and ion as the water polarizes in the electric field of the ion. This avoids overbinding and yields a model that can reproduce both the experimental thermodynamic and kinetic properties, a capability not attained in any other FF to our knowledge. It highlights the importance of explicitly including electronic polarizability in a FF as well as the advantage of using a Drude oscillator particle in representing the electronic degrees of freedom. It also captures the cation−π interactions for aromatic side chains more accurately.[518] As parametrization of ions and charged molecules is typically carried out to reproduce experimental data in the infinite dilution limit, accounting for the osmotic pressure made it possible to extend the models to concentrated solutions via pair-specific LJ (NBFIX) and Thole electric shielding (NBThole).[519] A similar philosophy was exploited to optimize the parametrization of amide solutions.[520] Early efforts led to a preliminary set of optimized parameters for ion−protein interactions,[521] although additional tests revealed a number of issues that are currently being addressed.

10.2. Proteins. Developments in the C36 additive protein FF have involved two iterations building on the C27 FF, also known as C22/CMAP.[473] The revisions primarily involved optimization of the bonded parameters with only minimal changes in non-bonded terms.[476] The first iteration published in 2012 yielding C36 focused on the CMAP term targeting NMR solution data for non-Gly, non-Pro amino acids, the CMAP terms for Gly and Pro residues targeting high-level QM data and the χ_1 and χ_2 side-chain dihedral parameters, with the latter targeting condensed phase data from simulations of $(Ala)_4$-X-$(Ala)_4$ model peptides.[522] Subsequent focus was on the CMAP term to account for oversampling of the α_L conformation with C36 and improvements in the interaction between Arg and carboxylate groups using an off-diagonal LJ term (NBFIX in CHARMM nomenclature), yielding C36m.[144] These additional optimizations lead to improved treatment of intrinsically disordered peptides (IDPs) while maintaining accurate treatment of folded proteins. C36m is considered the default FF for additive protein simulations.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

An interesting outcome of that study regarded the role of LJ interactions between the water molecule and the protein in the sampling of folded versus unfolded states of IDPs. While changing the magnitude of the water−protein interactions could improve the equilibrium between folded and unfolded states for a specific protein, a general solution that universally treats all IDPs in the context of an additive FF may not be accessible. Additional works included improved treatment of cation−π interactions,[523,524] and halogen−protein interactions important for ligand−protein simulations.[525] Furthermore, the additive FF was extended to over 100 non-standard amino acids[526] and to α-methyl amino acids.[527]

Advances in the Drude FF have been substantial as prior to 2009 only water, ion, and small molecule parameters had been published. Parameters have been released for many other biological molecules, including significant work on the protein portion of the FF. Small molecule developments included heteroaromatics,[528] sulfur-containing compounds,[529] and ethers[530] with the Drude FF shown to yield accurate hydration free energies facilitated by the use of atom-pair-specific LJ parameters (i.e., NBFIX).[531] Building on the foundation of the small molecule parameters the first generation of the Drude protein FF, Drude-2013, was presented. It overcame challenges with moving from individual molecules to a polymer in a polarizable FF, where unexpected overpolarization was avoided by accounting for the conformational properties of the polypeptide backbone.[532]

Application to polypeptide simulations revealed importance of explicit polarization in both peptide folding[533] and unfolding,[534] by capturing cooperativity inaccessible to additive FFs. A number of issues found in Drude-2013 including the stability of β-sheet structures, led to additional optimization yielding Drude-2019.[535] Improvements involve both the polypeptide backbone and side-chain conformational properties including optimization of the electrostatic parameters of the atoms linking side chains to the backbone, as well as the treatment of cation−π and anion−π interactions.[536] Drude-2019 shows systematic improvements as compared to C36m and Drude-2013, and it allows stable simulations of proteins on the microsecond time scale. While the Drude FF has largely been developed assuming an explicit solvent model, a Poisson−Boltzmann (PB) implicit solvation model has been developed[537] and subsequently used to predict pK_a values of proteins.[538] With the Drude PB model, pK_a's calculated for 8 proteins were insensitive to the assigned dielectric constant, in contrast to the need for a value of 4 with C36m. This indicates a potential advantage of the polarizable model in implicit solvent approaches.

10.3. Nucleic Acids. Both additive and polarizable FFs for nucleic acids have been updated. Updates to the C27 nucleic acid FF include improved treatment of RNA, largely to account for contributions of the 2′OH group to conformational heterogeneity[539] as reported in a combined QM/bioinformatics study.[540] Work on DNA focused on the equilibrium between the BI and BII conformations in duplex structures.[541] In both cases, adjustments to the C36 FF only involved select dihedral parameters, suggesting that minimal improvements within the non-polarizable additive approximation were necessary. Beyond canonical DNA and RNA, the C36 FF was extended to a range of naturally modified ribonucleotides as required for the ever increasing list of non-coding RNAs.[542]

Given the high charge density of polyanionic nucleic acids, a polarizable model is of particular interest. Development of the Drude FF was based on carbohydrate, ion, and heteroaromatic parameters. The first step involved optimization of Drude parameters for nucleic acid bases targeting a range of QM data for interactions with water and for base−base interactions, as well as experimental data including base crystal geometries and heats of sublimation.[543] They were then combined with initial parameters for the phosphodiester linkage yielding the Drude-2013 DNA model[544,545] that was iteratively optimized with particular emphasis on dihedral parameters associated with phosphodiester, sugar, and sugar−base glycosidic linkages. The optimization involved comparison with experiments including crystal data, to ensure suitability in simulations of duplexes in the condensed phase such as the equilibrium between A- and B-DNA and the BI/BII forms. This initial model improved agreement with experimental data regarding base flipping,[546] and yielded insights into the distribution and competition between ions around duplex DNA.[547] Impact of ions on DNA conformation including the minor groove width could also be addressed.[548] Such results cannot be captured well by additive FFs, again emphasizing the utility of the polarizable model for studying charged species.

Subsequent optimization of the Drude-2013 DNA FF focused on the underestimation of base stacking in duplexes and unwinding of Z-DNA. It involved additional QM calculations on Z conformations and application of higher-level model chemistries for other QM data.[549] The resulting model reproduced both crystal and solution scattering data over a range of duplexes in microsecond simulations.[550] The FF was also extended to RNA[551] which focused on the role of the 2′OH group on the conformational properties using QM data on RNA-specific model compounds. Condensed phase testing involved stem-loop structures, adenine riboswitch, and canonical duplexes, showing good agreement with crystallographic and NMR data.

The combination of the DNA and RNA FF, termed Drude-2017, was applied successfully to a number of systems including quadruplexes where the ions in the G tetramer are stabilized by the explicit inclusion of electronic polarizability.[552] However, a tendency of the Drude-2017 FF to overpolarize the Drude particle during MD simulations was noted. While this was addressed by using the Drude hardwall constraint[553] it represents a non-adiabatic condition. The electrostatic parameters were subsequently adjusted to yield a model that was successfully used in simulations of RNA hairpins.[554]

10.4. Lipids. Significant advances were made to the additive and polarizable lipid FFs. Revised parameters for 6 neutral lipids were introduced, yielding the C36 lipid FF.[555] As background, previous CHARMM lipid FF[555] required an applied surface tension to avoid unphysical bilayer surface area contraction in NPT simulations. Adjustments to charges and torsion angles in the headgroup region in C36: (1) reduced the surface tension to zero at the observed experimental surface area per lipid for free-standing dipalmitoylphosphatidylcholine (DPPC) bilayers; (2) increased area compressibility to experimental ranges; and (3) captured the experimentally observed splitting in deuterium order parameters for carbons in glycerol and carbon 2 of the chain 2. The C36 FF was further validated by the agreement with experimental bending constants[556] and spontaneous curvatures[557,558] (those for bilayers required new code for pressure profiles described in

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Section 12.3). Extensions to new lipids are ongoing. To date, common phospholipids are parametrized, including 13 variants of inositol lipids, sphingolipids, 5 hydroxylations for ceramide lipids, ether lipids, glycolipids, and acyl chain variants (saturated, monounsaturated, polyunsaturated, branched, and cyclic). Excluding the nearly unlimited variations in glycolipids and lipopolysaccharides, over 300 lipids have been parametrized for C36[559−561] and they are readily available in CHARMM-GUI. A united-atom representation wherein hydrogen atoms are combined with their bonded heavy atom, has also been formulated and tested for most common lipids (C36UAr).[562] It is currently being extended to other lipid head groups and chain types such as sphingolipids.[563]

Despite its extensive refinements and wide usage, C36 has two fundamental limitations: sensitivity of the truncation method used for the LJ interactions, and lack of polarizability. The former manifests as inconsistent bilayer and monolayer surface tensions, which is because bilayers are parametrized to agree with their experimental surface area at particular temperatures. The acyl chain−air interface of monolayers is highly sensitive to truncation of the LJ potentials, causing underestimation of surface tension when using the same parameters that otherwise yield accurate results for bilayers. Conversely, the surface tension of the C36 DPPC monolayer agrees well with the experimental value when long-range LJ terms are included, but the bilayer contracts.[420] It was thus necessary to parametrize the bilayer and monolayer consistently. While it can in principle be carried out with a truncated LJ potential, it is physically more reasonable to parametrize both with long-range interactions, which also avoids the sensitive dependence of bilayer properties such as for phase changes, on user-specified cutoff values. While a computationally efficient way of including long-range LJ terms in anisotropic systems such as bilayers and monolayers was unavailable when C36 was developed, subsequent incorporation of LJ-PME[421,422] (Section 8.5) led to reparameterizing C36 to C36/LJ-PME.[564,565] Consistency of bilayer and monolayer surface tensions for DPPC in C36/LJ-PME was obtained without compromising the overall quality of C36 for bilayers. Yet, monolayer isotherms at very large surface area where the surface tension of water−air is important are not well-described in C36/LJ-PME because the water−air surface tension of TIP3P (the default water in the additive CHARMM FF) is substantially lower than the experimental value.[417]

The second limitation of C36, the lack of polarizability, manifests as water permeability in saturated lipids being 5-fold lower than experimental values.[566] This is because the transfer free energy of water into hexadecane (a good model for the interior of a bilayer) is overestimated by 1 kcal/mol. Also, the dipole of the additive water cannot readjust when it is in the lipid environment. This motivated the development of the CHARMM Drude polarizable FF. Early versions of the lipid Drude FF[553,567,568] provided insight into membrane dipole potentials,[567] the mechanism of permeation of arginine as a function of membrane thickness,[569] as well as ion conduction along the narrow gramicidin A channel.[570] While these studies demonstrated the importance of a polarizable FF for membranes, the initial parametrization of phospholipid molecules had a number of shortcomings, including overestimated bilayer area compressibility. Furthermore, it was optimized with a truncated LJ potential without accounting for long-range dispersion. Bilayer surface areas and compressibility of the revised Drude-2023 FF[571] agree much better with experiments. More importantly, Drude-2023 yields more accurate dipole potentials, water permeability, monolayer isotherms, and lipid diffusion constants compared to C36 or C36/LJ-PME. Efforts toward a more comprehensive collection of lipids in the context of Drude FF, including charged lipids and ceramide-based lipids, are ongoing.

10.5. Carbohydrates. Building FF for carbohydrates poses a particular challenge given the wide range of monosaccharides, including both furanoses and pyranoses, the large number of chemical functional groups beyond hydroxyls, and various glycosidic linkages in poly- and oligosaccharides. Additive carbohydrate FF developments since 2009 included acyclic species[572] and furanoses[573] along with the required glycosidic linkage.[574] Significantly increasing the coverage of the FF was the inclusion of a variety of chemical groups along with testing on polysaccharides and glycan−protein interactions.[575,576] They together represent the carbohydrate portion of the C36 FF that has been widely used for carbohydrates, glycolipids and glycoproteins.

A similar path was taken with the Drude polarizable FF. Extensive non-bonded parameter optimization was undertaken on acyclic polyalcohols,[577] aldehydes, and ketones,[578] as required for the treatment of linear alcohols. FF for furanose[579] and pyranose[580] monosaccharides were completed and subsequent adjustments were made to the LJ parameters of pyranoses to improve the treatment of stacking interactions that led to better diffusion behaviors of glucose.[581] This was followed by parametrization of glycosidic linkages involving furanoses and pyranoses[582] and extension to N-acetyl groups[583] and both N- and O-linkages for glycoproteins.[584] Application of both the C36 and Drude FF to mannose disaccharides showed good agreement with NMR observables.[585] As with the rest of the Drude FF, the nomenclature, with few exceptions, has been designed to be identical to that of C36 for ready access.

10.6. Small Molecules. CGenFF and DGenFF have been developed to greatly broaden the coverage of FFs by rapidly generating topologies and parameters for a wide range of molecules including those of medicinal chemistry and ionic liquids. CGenFF initially leveraged the wide collection of topologies and parameters of C36. Its coverage then extended to drug-like molecules by applying an optimization protocol that maintains compatibility with C36. While the initial CGenFF paper focused on the general philosophy of the model and details of parameter optimization,[145] the machinery for rapid application to small molecules was already in place. This included bond perception and atom typing algorithms along with the charge assignment protocol compatible with C36.[17,18] In addition, the CGenFF program outputs penalties associated with charges and parameters not in the existing CGenFF parameter set, where the penalty is assessed based on the similarity between the algorithmically derived parameter and those available in the FF. As the penalty is not a direct measure of the "quality" of a given parameter, in many cases, parameters with relatively high penalties are often appropriate for modeling and simulation. CGenFF has been extended to include sulfonyl- and halogen-containing compounds.[586,587] Treatment of halogens made use of lone pairs on aromatic Cl, Br and I atoms, allowing for modeling of halogen bonds involving weak favorable interactions with hydrogen bond acceptors along the C−X bond. Another major extension included parametrization of non-standard amino acids

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

mentioned above,[526] which were treated with CGenFF combined with the C36 protein FF.

The small-molecule DGenFF was designed to be analogous to CGenFF with some important differences. Notably, DGenFF takes advantage of the original CGenFF program in which the bond perception, atom typing, parameter assignment and bonded penalty assignment algorithms were based on a rule-based approach that creates a rules file specific for the DGenFF while using the same CGenFF program. Assigning electrostatic parameters for the Drude FF requires partial atomic charges, atomic polarizabilities, and Thole scaling factors, where a deep neural network (DNN) was developed for each term,[588] building upon an earlier DNN model.[589] Features were based on atom connectivity up to 1−5 bonded atoms along with local through-space atom type−atom type pairs with training targeting QM data on nearly 40,000 small model compounds (<200 Da). This approach rapidly assigns electrostatic parameters along with penalties based on populations of different atom types and their connectivity in the DNN training set. Additional validation against QM dipole moments and molecular polarizabilities on 900 FDA-approved compounds showed excellent agreement, indicating that the method is appropriate for drug-like molecules 200 to 700 Da in size. Note that including distance features in the DNN to model asymmetric electrostatic parameters on lone pairs requires that molecules have approximately correct 3D geometries as well as correct ionization and tautomer states.

Efforts are ongoing to extend the coverage of DGenFF comparable to that of CGenFF. To date, extension to halogens has been completed.[590] It includes the presence of lone pairs to accurately treat halogen bonds along with careful optimization of the anisotropic atomic polarizability on the halogens Cl, Br, and I and inclusion of LJ parameters on Drude particles, as was done for water−Mg^{2+} interactions discussed above. These latter terms allow for accurate modeling of out-of-plane interactions with hydrogen bond donors where the halogen atom serves as a hydrogen bond acceptor, which is more favorable than halogen bonds and is present in a large number of ligand−protein complexes.[591] Further extension of DGenFF will be facilitated by a DNN-based workflow to optimize LJ parameters of new atom types.[592] Upon completing full coverage, global optimization of LJ parameters will be undertaken, as done for a subset of atom types in the Drude FF.[593] The resulting FF parameters yield good agreement with both pure solvent properties and hydration free energies for a large collection of small molecules. To our knowledge, agreement with both classes of condensed phase properties has not been attained with any additive FF, even when a similar global optimization protocol was used.[594]

A final issue concerning FFs in general is the validity of implementation.[595] CGenFF is based on a specific algorithm to assign atom types yielding model compound topologies used for parameter assignment and optimization. When other algorithms are used based on analogy, the resulting charges and parameters are inconsistent with those on which the FF was optimized. For correct implementation of CGenFF, individual molecules can be uploaded online by users from educational institutions at https://cgenff.silcsbio.com/ and the CGenFF program can be obtained at no charge for users from educational institutions from Silcsbio LLC (silcsbio.com).

11. MIXED QUANTUM MECHANICS/MOLECULAR MECHANICS (QM/MM) METHODS

11.1. Background. QM/MM methods are practical and efficient approaches for simulating chemical reactions in condensed phase including enzyme catalysis.[596−599] A QM method is necessary for modeling changes in electronic structure such as bond formation and cleavage, photochemical reactions, and electron transfer in redox catalysis by metalloenzymes. However, it is neither practical nor necessary to treat an entire substrate−enzyme complex and the surrounding solvent at the QM level. This is further complicated by the need to sample multiple protein and solvent configurations to determine the free energy change along a reaction pathway. A combined QM/MM approach addresses this challenge by selectively applying the QM treatment to a region of the system involved in the reaction, such as the substrate, cofactors and key amino acid residues directly participating in the chemical event. This 'QM subsystem' is embedded in the rest of the system represented by an MM FF.[596,600,601] Because of its effectiveness and simplicity, QM/MM methods have become the *de facto* choice for simulating enzyme reactions.[597−599,602,603]

A QM/MM approach was first implemented in CHARMM by Field and Bash in 1987,[601,604] employing the SE 'neglect of diatomic differential overlap' (NDDO) method along with the CHARMM FF. QM/MM approaches have since continuously embraced diverse methods, including *ab initio* (AI), DFT, and SE QM alternatives. The SE-QM methods encompass both the NDDO-based models[161,605−608] and the density functional tight binding (DFTB) methods, also referred to as the self-consistent-charge DFTB (SCC-DFTB) methods.[609] Both of them are incorporated into CHARMM.

The QUANTUM module was the first SE-QM/MM method implemented in CHARMM, which was based on the MOPAC program (version 4.0).[610] Subsequently, two new NDDO-based SQUANTM[3] and MNDO97[147,611] modules were added. The SQUANTM module was based on an implementation in the AMBER program,[612] and the MNDO97 module was derived from a stand-alone MNDO97 program.[613] The latter has recently been rewritten for computational speed and parallelization.[147] The DFTB method was similarly implemented in CHARMM.[163] Due to their computational efficiency, these SE-QM/MM methods are frequently used in conjunction with other free energy simulation techniques, including US,[138] SM,[139,614] reaction path,[148,615] and FEP.[604]

For AI and DFT-based QM/MM methods, CHARMM provides robust interfaces to external softwares including Q-Chem,[616] GAMESS-US,[617,618] GAMESS-UK,[619] CADPAC,[620] and Gaussian16.[621] In addition, the MSCALE module provides a flexible means for accessing other QM programs, such as MOLPRO.[622] Unlike the SE-QM/MM modules of CHARMM, other packages for AI and DFT calculations must be obtained separately. Except for GAMESS-US and GAMESS-UK that can be compiled as a single executable within CHARMM, other packages should be installed separately.

11.2. QM/MM Potentials and Practical Considerations. The effective QM/MM Hamiltonian operator is

$$\hat{H}_{\text{eff}} = \hat{H}_0 + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \tag{51}$$

where $\hat{H}_0$ and $\hat{H}_{\text{MM}}$ describe the QM and MM subsystems, respectively, and $\hat{H}_{\text{QM/MM}}$ describes the interaction between

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

the two. The latter is further decomposed into electrostatic, van der Waals, and the QM−MM boundary terms:

$$\hat{H}_{QM/MM} = \hat{H}^{elec}_{QM/MM} + \hat{H}^{vdW}_{QM/MM} + \hat{H}^{boundary}_{MM} \quad (52)$$

In CHARMM, $\hat{H}^{elec}_{QM/MM}$ is solved self-consistently with $\hat{H}_0$, while $\hat{H}^{vdW}_{QM/MM}$ is modeled with a LJ potential. The total energy of the system is

$$E_{tot} = \langle \Psi_{el}|\hat{H}^0 + \hat{H}^{elec}_{QM/MM}|\Psi_{el}\rangle + E^{vdW}_{QM/MM} + E^{boundary}_{MM} + E_{MM} \quad (53)$$

where Ψ_{el} represents the Hartree−Fock wave function for electrons and nuclei of the QM region. For evaluation. the first three terms in Eq. 53 are determined in the QM/MM module while E_{MM} uses MM energy routines. For performing MD simulation or energy minimization, the QM electron density matrix from the previous MD/minimization step can be used as the initial guess for the next SCF calculation in Eq. 53, with optional addition of small random perturbations to reduce hysteresis and accelerate the convergence of the QM and QM/MM energies. This is implemented in all SE-QM/MM modules as well as in AI-QM/MM modules supporting GAMESS-US, GAMESS-UK and Q-Chem.

In Eq. 53, $E^{boundary}_{MM}$ addresses cases where the QM and MM division occurs across covalent bonds leaving the QM region with unsaturated dangling bonds. This commonly occurs in enzyme simulations where specific side chains are included in the QM region, for which three methods are available in CHARMM. The first is the hydrogen link (H-link) atom approach, where a hydrogen atom is added within the QM region to saturate and cap the dangling bond.[601,623,624] The H-link method is conceptually simple, so it has been widely adopted in various packages. In CHARMM, users can introduce an MM angle term to keep the H-link atom aligned with the replaced QM−MM bond. Also, charges on nearby MM atoms can be reassigned to minimize artificial polarization around the QM−MM boundary. The second model is the double-link atom method (DLAM) where an additional H-link atom is introduced at the MM atom site of the QM−MM covalent bond to achieve bond saturation at both of the loose QM and MM ends[625] (Section 11.6). This method can be used together with delocalized Gaussian MM (DGMM) charges to mimic the delocalization of charge densities on MM atoms.[626] The third is the generalized hybrid orbital (GHO) method.[627−629] While the first two methods introduce additional degrees of freedom via the link atoms and alter local electrostatic potential by adjusting partial charges, the GHO method treats the QM boundary atom as a special *sp*3-hybridized carbon and also as an MM atom connected to nearby MM atoms. Its three *sp*3-hybrid orbitals pointing toward the connected MM atoms called auxiliary orbitals, are fixed with their electron densities assigned based on their MM charges. The remaining *sp*3-hybrid orbital called the active orbital, is optimized during the SCF iteration. In addition, MM FFs are applied to the GHO atoms to maintain the surrounding geometry.

The above boundary methods, particularly the GHO method, are specifically designed for covalent bonds between two *sp*3-hybridized carbon atoms, and they are not recommended for arbitrary covalent boundaries. This ensures that the covalent boundary does not perturb the geometry around the QM−MM bond and minimizes artificial polarization of the QM electron density. The GHO method implemented in CHARMM supports all four SE-QM/MM modules (QUANTUM, SQUANTM, MNDO97 and SCC-DFTB) and AI/DFT QM/MM methods through the GAMESS-US interface. The H-link atom approach is available for all QM/MM modules of CHARMM.

In QM/MM simulations, one must decide on: (1) the QM model, (2) atoms for the QM region, and (3) representation for the QM/MM covalent boundary, and (4) the boundary condition of the whole system.[630,631] For (1), among AI-QM, DFT and SE-QM, computational errors contributing to the final results should be considered. While AI-QM and DFT methods are more accurate, their high computational cost limits routine use in extensive MD and free energy simulations. Thus, selection of the QM theory level and the basis set should be tailored to individual problems. Also, the high accuracy of computationally expensive QM methods such as a coupled-cluster model may be overshadowed by the statistical noise itself. In such cases, an SE-QM method would be more suitable for lengthy QM/MM simulations. However, they require calibration against AI-QM/DFT levels for the reaction under consideration.[608,632−634] Among the three NDDO-based SE modules, QUANTUM and MNDO97 in CHARMM now have the option to read non-standard parameters without modifying the source code, obviating the need to rebuild the executable for reaction-specific parametrization.

About the choice of the QM region, there are recent debates about the minimum QM region size required for convergence.[377,635−640] It affects the computational cost and the extent of sampling needed. Advances in efficient algorithms and specialized hardware enable systematic exploration of the QM size for the desired accuracy. However, such an investigation is currently feasible only for relatively small systems, and chemical intuition still remains crucial for selecting the QM region. In any event, we note that QM/MM methods are fundamentally empirical approximations. There is no reason to expect that an arbitrary combination of QM and MM models will reproduce the full quantum results. One should carefully optimize parameters for separating a full QM system into two distinct QM and MM models, to determine the minimum size of the QM region.

CHARMM can perform the QM/MM calculations using both PBC and the solvent boundary condition.[630] When using PBC, the QM/MM-Ewald[641] and QM/MM-PME methods[612,642] can be used. In this case, the E_{MM} term in Eq. 53 includes MM−MM interactions with all images as for the regular PME and the $\hat{H}^{elec}_{QM/MM}$ term includes long-range electrostatic interactions of all MM and QM periodic images with the QM charges. The QM/MM-Ewald method is available in all SE-QM/MM methods, while the QM/MM-PME method is available in the SQUANTM and MNDO97 QM modules; the QM/MM-Ewald method is also supported by the AI/DFT-QM/MM method employing the QChem package.[643] For the DFTB method, the GSBP method is also available as an alternative way to incorporate long-range electrostatic interactions into the QM/MM framework.[644,645] Otherwise, it is recommended not to use any cutoff scheme for non-bonded interactions for balanced interactions between the QM−MM and MM−MM pairs.[641] This is because in Eq. 53, any MM atom included in the QM/MM interactions, e.g., those within the cutoff distance of any QM atom, interacts

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

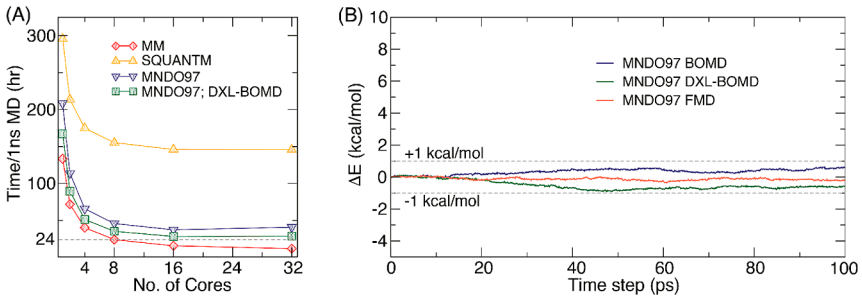


Figure 14. Performance of QM/MM methods.[147] (A) Wall time in hours per 1-ns MD simulation for insulin receptor kinase versus the number of CPU cores, for the MM-only, SQUANTM, MNDO97 BOMD, and MNDO97 with DXL-BOMD methods. (B) Energy conservation in different SCF accelerator implementations for adenylate kinase, showing less than 1 kcal/mol deviation of the total energy during 100-ps NVE MD simulations. Simulation systems consist of (A) 76 QM atoms and 28,823 MM atoms and (B) 92 QM atoms and 47,201 MM atoms. In both cases, the QM region is treated with the AM1/d-PhoT SE QM model[608] for the MNDO97 module and with the AM1 model for the SQUANTM module.

with all QM atoms and thus directly polarizes the QM electron density.

11.3. Recent Advances in SE-QM/MM Methods. Main strengths of SE-QM/MM methods over AI/DFT-QM/MM methods are their efficiency and flexible functional forms that allow recalibration against target data.[608,633,634] Extended MD simulations within reasonable computational time is thus possible while yielding accuracy tailored to individual systems. In enzyme mechanism studies, SE-QM/MM methods are frequently employed together with the RXNCOR module for US simulations and more recently, with the SM.[139,646,647] They can also be used in FEP simulations such as calculating solvation free energies of solutes and ligand pK_a values, functionalities available in the QUANTUM, SQUANTM, and SCC-DFTB modules.[598,645,648,649]

CHARMM also has several acceleration algorithms for the SE-QM/MM methods, including the direct inversion of the iterative subspace (DIIS) extrapolation scheme for faster SCF convergence[650,651] and the pseudodiagonalization algorithm for the Fock matrix. In addition, the MNDO97 module has recently incorporated the MPI parallelization and several new SCF accelerators, achieving more than 10-fold speed up (Figure 14).[147,165] The newly implemented SCF accelerators are as follows:

11.3.1. Extended Lagrangian MD (ELMD).[165] This method[652] performs MD simulations with the electron density of the QM subsystem propagated by the Lagrangian:

$$\mathcal{L}_{\mathrm{ELMD}} = \frac{1}{2}\sum_{A} M_A \dot{\mathbf{R}}_A^{\,2} + \frac{1}{2}\sum_{i,j} m_{i,j}\dot{P}_{ij}^{\,2} - E_{\mathrm{tot}}(\mathbf{R}, \mathbf{P}) - \mathrm{Tr}[\mathbf{\Lambda}(\mathbf{P}^2 - \mathbf{P})] \quad (54)$$

Here, $\mathbf{R}$ denotes the QM and MM coordinate with mass M_A (A: atom index). $\mathbf{P}$ is the electron density matrix with mass m_{ij} for the corresponding element. $\dot{\mathbf{R}}_A$ and $\dot{P}_{ij}$ are time derivatives. The Lagrange multiplier $\mathbf{\Lambda}$ enforces the idempotency constraint on $\mathbf{P}$. Alternatively, $\mathbf{P}$ can be propagated using the curvy-steps unitary update algorithm.[653] Both methods avoid time-consuming SCF iteration at the expense of a smaller integration time step for the density matrix propagation. This limitation can be alleviated by applying the multiple time step (MTS) approach where nuclear coordinates are propagated with a larger time step, typically 0.5 or 1 fs, as commonly used in QM/MM MD simulations.

11.3.2. Extended Lagrangian Born–Oppenheimer MD with Dissipation (DXL-BOMD).[147] The idempotency condition of Eq. 54 is modified into a harmonic restraint for the auxiliary density variable $\mathbf{D}$, to oscillate around the converged SCF density:

$$\mathcal{L}_{\mathrm{DXL\text{-}BOMD}} = \frac{1}{2}\sum_{A} M_A \dot{\mathbf{R}}_A^{\,2} + \frac{m_D}{2}\sum_{i,j} \dot{D}_{ij}^{\,2} - E_{\mathrm{tot}}(\mathbf{R}, \mathbf{P}) - \frac{\kappa_D}{2}(\mathbf{D} - \mathbf{P})^2 \quad (55)$$

The true (SCF) density $\mathbf{P}$ is approximated by $\mathbf{D}$ that serves as the initial guess in the SCF iteration, and κ_D is the force constant. $\mathbf{D}$ is extrapolated based on a predetermined number of previous SCF densities.[654,655] The SCF iteration is then performed for a given number of SCF steps. In practice, Eq. 55 is solved in the limit of vanishing m_D, resulting in coupled equations of motion, one for the nuclear position and the other for $\mathbf{D}$, thereby eliminating the dependence of results on m_D. This method is also supported by the DFTB QM/MM module.[655]

11.3.3. Fock Matrix Dynamics (FMD).[147] The Fock matrix is directly extrapolated based on its elements determined from previous MD steps, followed by regular SCF iteration until convergence.[656,657] Both DXL-BOMD and FMD methods significantly reduce the number of SCF iterations compared to conventional (BOMD) SCF calculations while maintaining energy conservation.[147]

In addition to efficiency, SE-QM/MM methods in CHARMM are being developed to improve accuracy. Recognizing the importance of non-bonded interactions, the MNDO97 and DFTB modules have implemented Grimme's dispersion and hydrogen bond correction terms.[638,658,659] To further improve the quality of the PES, a simple valence bond-like (SVB) term has been introduced in the NDDO-based SE-QM/MM modules.[646,660] In addition, the SQUANTM module integrates the SE-QM/MM and GAMESS-UK AI/DFT-QM/MM methods, introducing a dual-level approach that interpolates the QM/MM energy to the AI/DFT-QM/MM level of theory.[642] This method is compatible with the MTS algorithm so that MD simulations are performed at the SE-QM/MM level while simultaneously correcting energies and gradients at the target AI/DFT-QM/MM level over a longer time step (Figure 15). These developments enable simulations of highly challenging systems with unprecedented accuracy and efficiency, pushing the boundaries of QM/MM methods.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

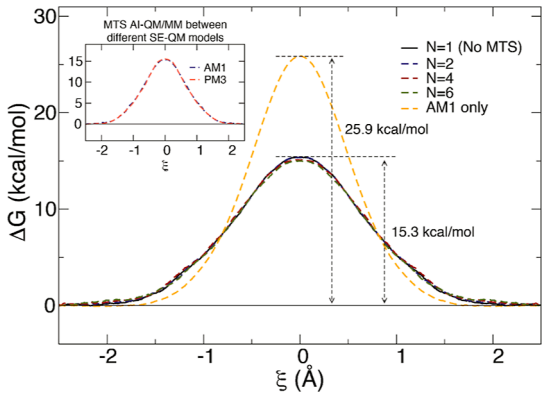


Figure 15. PMF for the S_N2 reaction between CH_3Cl and Cl^- in water.[642] Results from MTS simulations with varying number of MD steps *N*, for evaluating the AI-QM/MM correction term. The MTS AI-QM/MM simulations were carried out using the AM1 and HF/3-21G methods for the low- and high-level QM theories, respectively. "AM1 only" results are from the AM1 SE-QM/MM simulations. The inset compares the impact of low-level QM theory on the PMF, while the high-level theory remains at the HF/3-21G level.

11.4. Path-Integral-Free Energy Perturbation for Nuclear Quantum Effects. An important approach for treating nuclear quantum effects (NQE) is the Feynman path integral (PI) formalism[661] that describes the wave function by an ensemble of paths weighed by the classical action for each, to capture Schrödinger's delocalized wave behavior. It is readily generalizable to multiparticle systems, and it naturally accounts for thermal effects and sampling can be performed in MC and MD simulations.

The quantum transition state theory (QTST) rate constant may be computed with PI (PI-QTST).[662] Consider a system composed of a set of QM atoms with coordinates **r** embedded in a bath of classical atoms with coordinates **R** in thermal equilibrium. The QM partition function Q_Q can be written as the trace of the thermal density matrix ρ

$$Q_Q = \mathrm{Tr}(\rho) = \int d\mathbf{R} \int d\mathbf{r} \rho(\mathbf{r}, \mathbf{r}; \mathbf{R}, \beta) = \int d\mathbf{R} \int \Pi_{j=1}^{P} d\mathbf{r}_j \rho(\mathbf{r}_j, \mathbf{r}_{j+1}; \mathbf{R}, \tau) \quad (56)$$

where *P* is the number of quasi-particles or beads and $\tau = \beta/P$ ($\beta = 1/k_BT$). For a closed ring-polymer chain, $\mathbf{r}_1 = \mathbf{r}_{P+1}$. In the high-*T* limit ($\tau \to 0$ and $P \to \infty$), the semiclassical primitive approximation[663] can be used for ρ

$$\rho_{PA}(\mathbf{r}_j, \mathbf{r}_{j+1}; \mathbf{R}, \tau) = \left(\frac{m}{2\pi\tau\hbar^2}\right)^{D/2} \times e^{-[(m/2\tau\hbar^2)(\mathbf{r}_{j+1}-\mathbf{r}_j)^2 + (\tau/2)\{V(\mathbf{r}_{j+1};\mathbf{R}) + V(\mathbf{r}_j;\mathbf{R})\}]} \quad (57)$$

where *V* is the system potential (i.e., a QM/MM potential), *m* is the quantum particle mass, and *D* is the dimension of **r**. The description above is isomorphic to a classical ring of beads system connected via harmonic springs, and forms the basis for the implementation in CHARMM where the bead distribution is sampled using MC simulations.

The rate constant of PI-QTST[662] is defined as

$$k_Q^{\text{PI-QTST}} = \Omega_{FP} Q_Q^{\ddagger} / Q_Q^{RS} \quad (58)$$

where Ω_{FP} is a free-particle (FP) prefactor, and $Q_Q^{\ddagger}$ and Q_Q^{RS} are the quantum partition functions for the transition and reactant states, respectively. Although one can compute the PI-QTST directly, when using expensive potentials like QM/MM, it is convenient to compute the correction to the classical TST due to NQE

$$\delta^{\ddagger} = \frac{Q_Q^{\ddagger}/Q_C^{\ddagger}}{Q_Q^{RS}/Q_C^{RS}} \quad (59)$$

where the subscript *C* denotes the corresponding classical partition functions. The above includes both quantum vibration (zero-point energy) and tunneling effects. To calculate the quantum to classical ratio of the partition function, a double average scheme[664,665] can be used

$$\delta = \frac{Q_Q}{Q_C} = \langle\langle e^{-\tau \sum_{j=1}^{P} [V(\mathbf{r}_j;\mathbf{R}) - V(\mathbf{r}_c;\mathbf{R})]} \rangle_{FP,\mathbf{r}_c} \rangle_{V(\mathbf{r}_c;\mathbf{R})} \quad (60)$$

where $\mathbf{r}_c$ is the centroid coordinate. The outer (classical) average is obtained using standard simulation techniques, while the delocalized QM description comes from the inner FP average. This approach is practical since the classical and quantum simulations are performed separately. In CHARMM, the NQE atoms are defined as QM atoms. Currently, the PI method works with all SE-QM/MM modules in CHARMM as well as with AI/DFT-QM/MM using CHARMM and Q-Chem.

A well-known challenge with PI simulations is the difficulty with sampling the polymer ring due to the harmonic coupling between the beads (Eq. 57).[663] In CHARMM, the inner average in Eq. 60 for the FP PI sampling is performed using the bisection algorithm[666,667] extended to a ring of quasi-particles[668,669] or using the staging algorithm.[670,671] Since each new configuration in the bisection or staging PI sampling is independent of the distribution of beads in previous configurations, the PI rapidly converges, which is essential for accurate calculation of the NQE and absolute rate constants.[633,672−674] However, specialized techniques are required to precisely compute isotope effects due to minute differences in free energy. In CHARMM, a novel mass-perturbation technique termed PI-FEP, was developed to directly compute the free energy difference between isotopes.[675] Since relative free energies between the distributions of heavy and light particles are determined by FEP (Figure 16), the precision of the computed kinetic isotope effect (KIE) is of experimental quality for both the primary and secondary KIEs.[671,676−683]

To reduce the computational cost of PI simulations in CHARMM, higher-order factorization of the density matrix operator has been adopted,[684−686] which converges with a considerably smaller number of beads at the expense of computing the potential gradient in addition to the potential.[671] This method can be combined with the PI-FEP approach to efficiently compute the KIE.[679]

Information about tunneling can be obtained by inspecting the particle momentum distribution computed using open-chain PI (i.e., $\mathbf{r}_1 \neq \mathbf{r}_{P+1}$).[687−689] Whereas closed-chain PI only samples diagonal elements of ρ, open-chain PI simulations also sample off-diagonal elements, and it can sample both the anisotropic and isotropic momentum distribution of a transferring hydrogen ($H^+/H^-/H\cdot$) during a reaction.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

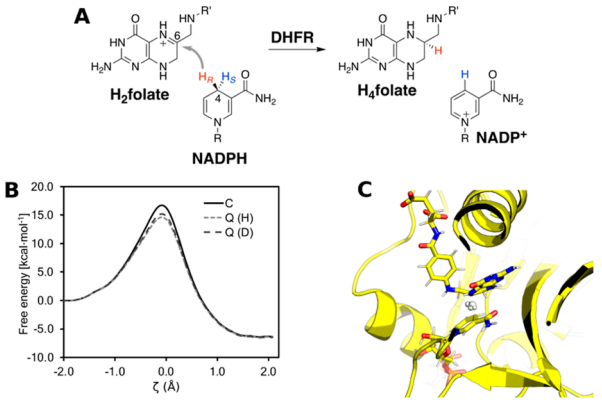


Figure 16. Hydride transfer reaction in dihydrofolate reductase.[633] (A) Reaction mechanism. (B) Classic (solid line) and quantum (dashed lines) PMFs for hydride and deuteride transfer. (C) Active site for the hydride transfer from NADPH to H_2 folate. The transferring hydride is described using PI with 32 beads.

11.5. Density Functional Tight Binding (DFTB) Module. One versatile QM/MM module in CHARMM is based on the DFTB approach,[692,693] which is an approximation method that aims to strike the balance between computational efficiency and accuracy. Also referred to as SCC-DFTB,[609] it is an SE QM method in the sense that it employs a minimal basis and the two-center approximation for electron integrals, which are key approximations for NDDO-based SE methods[694] such as AM1, PM3 and OM2 (Section 11.1). On the other hand, most of the parameters in the DFTB approach are computed based on atomic or diatomic molecules, and the most empirical aspect of the parametrization concerns those used to derive the atomic/diatomic electronic properties and the pairwise repulsive potentials (e.g., confinement radius). The most popular approach for biomolecular applications is the DFTB3/3OB model,[164,695] which has been parametrized for elements commonly encountered in organic and biomolecular systems: O, N, C, H, S, P, Na, K, Mg, Ca, Zn, Cu and the halogens. For recent reviews of the development and application of the DFTB3 method for condensed phase applications, see refs 658 and 696.

The DFTB3 model is integrated with MM model in CHARMM through the standard electrostatic embedding scheme[163] where the DFTB3 atoms are represented as Mulliken charges; alternative DFTB3/MM electrostatic interaction models have also been implemented[697] that consider the finite spatial distributions of the DFTB3 and MM charges. In terms of boundary conditions, the DFTB3/MM model can be used together with either a GSBP[644] or the PBC with either Ewald summation[649] or PME[641] (Section 11.2). For localized reactions, the DFTB3/MM-GSBP approach is computationally most efficient and generally agrees with the more expensive DFTB3/MM-PME approach.[141] For systems in which the chemical reaction is coupled with considerable conformational rearrangements, the PBC-based approach is more appropriate. Another technical detail relevant to many QM/MM applications is the flexible inner region ensemble separator (FIRES) potential[698] available in CHARMM (Section 11.7), which prevents the exchange of QM and MM water molecules and thus particularly important for solution reactions[699] or for solvent-accessible active sites.[700]

The DFTB3/MM approach can be used together with many key functionalities in CHARMM, especially various types of free energy simulations that are essential to quantitative analysis of chemical transformations. For chemical reactions, they include US with the RXNCOR module, an interface with PLUMED[701] for various metadynamics simulations,[682,690,702−704] and the SM available in the STRINGM module[139,297,705] (Section 7.3). Another useful approach for improved sampling is replica exchange US through the REPDSTR module.[297] For alchemical free energy simulations, DFTB3/MM works with the PERT module.[297,706] For applications such as redox potential[310] and pK_a calculations,[649,707] the DFTB3/MM model also works with the BLOCK module.

In the following, the DFTB3/MM method is illustrated with two types of free energy simulations. For chemical reaction, the catalysis in Usb1, an exoribonuclease that shortens the oligouridine tail of U6 snRNA,[690] is used as an example. In the proposed catalytic mechanism, two active-site histidine residues serve as the catalytic base and acid, respectively (Figure 17A). It was studied with DFTB3/MM metadynamics

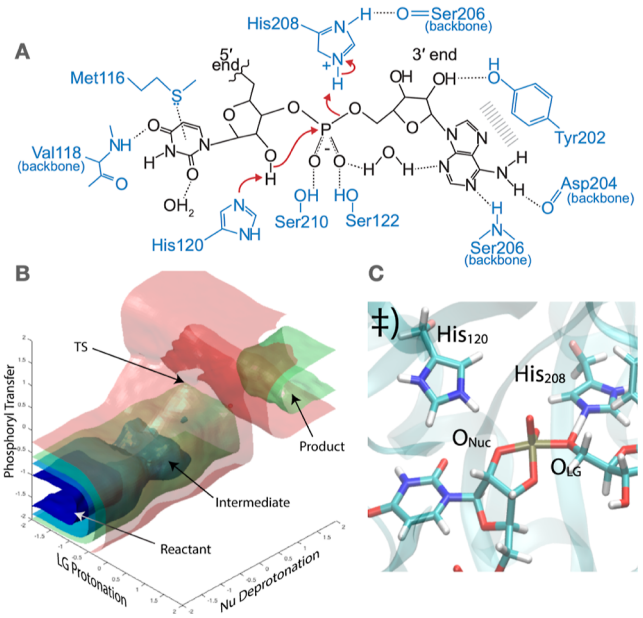


Figure 17. DFTB3/MM free energy simulations for the catalysis in Usb1.[690] (A) The putative catalytic mechanism of Usb1. (B) The 3D free energy surface from DFTB3/MM multiwalker metadynamics simulations. The three CVs describe proton transfers associated with the catalytic acid/base and the phosphoryl transfer (red arrows in panel A). (C) The transition state structure from the DFTB3/MM free energy simulations features the transfer of a single proton between H208 and the leaving group. This was subsequently confirmed with proton inventory experiments.[690] Panels A and C were adapted from ref 690, which was published by the Oxford University Press.

simulations using three CVs that describe proton transfer involving the catalytic base (H120), the phosphoryl transfer reaction, and the proton transfer involving the catalytic acid (H208), respectively. Multiwalker metadynamics calculations were run with 300−500 walkers, each of which was sampled for 0.5−1 ns, leading to a cumulative sampling of 0.1−0.2 μs for constructing the 3D PMF (Figure 17B). This level of sampling for QM/MM simulations is possible only with SE-type QM methods such as DFTB3, highlighting the value of calibrated SE QM/MM simulations for complex biomolecular

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

processes.[696,703] The transition state structure captured in the DFTB3/MM simulations (Figure 17C) suggests that one proton is in flight, and the predicted feature was subsequently confirmed experimentally.[690]

For alchemical free energy simulations, the binding selectivity of Mg^{2+} and Ca^{2+} in the Ca^{2+} binding protein carp parvalbumin (CP) and its mutant (D51A/E101D/F102W)[691] is considered (Figure 18A). Experimentally,[708] the WT CP was

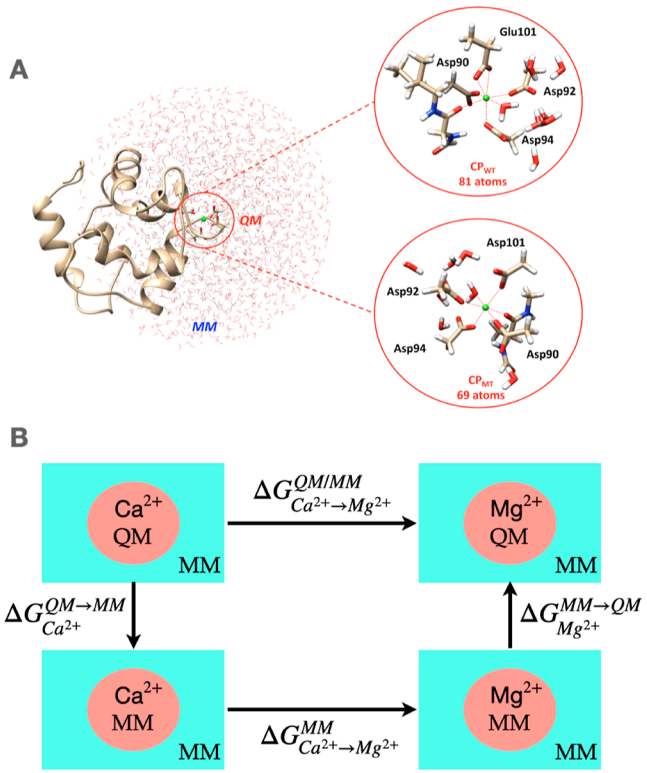


Figure 18. DFTB3/MM FEP simulations for the Ca^{2+}/Mg^{2+} binding selectivity in the Ca^{2+} binding protein, carp parvalbumin (CP), and its mutant (D51A/E101D/F102W). (A) The DFTB3/MM-GSBP setup that illustrates the QM regions for the WT and mutant CP simulations. (B) Thermodynamic cycle used to probe the free energy change for the Ca^{2+}/Mg^{2+} conversion in a given environment. In simulations, the horizontal transitions occur at the MM level, and in the vertical transitions, a metal binding site is converted between MM and QM treatments using the PERT module in CHARMM. Reproduced from ref 691. Copyright [2024] American Chemical Society.

measured to bind more strongly to Ca^{2+} by 5.6 kcal/mol; in the mutant, Ca^{2+} is still preferred over Mg^{2+}, although the selectivity is reduced to 1.6 kcal/mol. With a standard FF, the relative binding free energy of Mg^{2+} and Ca^{2+} to a protein is readily computed using alchemical free energy simulations; the metal ions are interconverted twice: once in the binding pocket and once in solution, and their free energy difference is the relative binding affinity. With non-polarizable FFs, such calculations did not yield the correct trend and the smaller Mg^{2+} was predicted to bind more strongly.[691,709] While the same set of alchemical free energy simulations can, in principle, be carried out at the DFTB3/MM level, we adopt an alternative thermodynamic cycle, which involves converting the description of the metal ion and its ligands between MM and DFTB3 levels[297] (Figure 18B). This has the advantage that structural changes during the MM/DFTB3 conversion are expected to be small and therefore convergence of the free energy simulation is rapid, minimizing the required DFTB3/MM computations. Encouragingly, with DFTB3/MM simulations,[691] the calculated $\Delta\Delta G_{bind}$ was ~6.2 kcal/mol, in good agreement with the experimental value. These results highlight the value of a QM description of the metal binding site and support the role of electronic polarization[709] and charge transfer[710] in metal binding to proteins. For the mutant, different binding site models led to considerable variations in the computed relative binding affinities. With a coordination number of seven for Ca^{2+}, which was shown by DFTB3/MM metadynamics to be the dominant coordination number for the mutant, the calculated relative binding affinity was ~4.2 kcal/mol, also in fair agreement with the experimental value.

11.6. Double Link Atom Method (DLAM). Accurately modeling chemical reactions in condensed phases using QM/MM is challenging, especially when partitioning across a covalent bond. Introducing dummy or link atoms serves as a bridge across the severed bond, acting as both a connection between QM and MM interfaces and an electron density cap for partitions. Typically hydrogen, a link atom can resemble the electronic character or features lost during truncation and is attached to a host group (HG) via the host atom (HA). The link atom is subject to interface with the MM and QM HGs. In the standard single link atom (SLA) scheme,[601,623] the link atom is added to the QM HA to compensate or neutralize the charge of the QM fragment and cap the QM Hamiltonian. However, the addition of the QM link atom introduces a number of challenges at the QM/MM interface, primarily the treatment of electrostatics.

Various approaches have been developed to address QM/MM boundary effects in the QM fragment by adjusting the magnitude of polarization from the MM fragment, including the excluded group (EXGR) scheme,[623] the charge shift scheme (CHSH),[711,712] the divided frontier charge (DIV) scheme,[624] and the distributed Gaussian (DG) method.[713,714] For example, the SLA scheme treats MM atoms as point charges and excludes the MM HA from the QM/MM electrostatics, leaving an artificial partial charge at the interface on the MM HG. The EXGR scheme corrects for the added unrealistic partial charge at the interface by excluding all partial charges of the MM HG, whereas the DIV scheme corrects for the unrealistic charge at the interface by redistributing the excluded MM HA partial charge to the MM HG, and the CHSH scheme introduces a dipole to counterbalance for the charge shift. Alternatively, the DG method includes all electrostatic interactions of the MM HG; however, the MM partial charges are represented as Gaussian charge distributions and utilize a smoothing potential or blur width (σ) to smear the MM electron density where optimal σ values vary depending on the physical property of interest.[626,714] Although these SLA-based schemes provide a balance for the QM fragment and interfacial electrostatics, they neglect the MM fragment which results in unbalanced forces and unrealistic electrostatics.

DLAM[626] addresses the above issues by the addition of a link atom to cap the MM fragment but it has been infrequently used due to its challenging implementation. It can now be called directly in CHARMM to add link atoms to both the QM and MM host fragments. The DLAMadd command adds a QM link atom "QQ" to the QM HA, similar to the single link atom command (ADDLink), and adds an MM link atom "QM" to the MM HA. By default, the link atoms are placed 1.0

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Å colinearly from the respective HA. The MM link, typically an MM hydrogen, bears a small partial charge that should preserve both the net charge and dipole of the MM HG. Balance can be achieved by shifting charge between the MM link and MM HA, but can be less straightforward in some systems. DLAM is used with the DG method and employs the same σ value for all MM atoms. As σ and MM link partial charge are free parameters in DLAM that can be tuned to balance electrostatics, identifying reliable parameters for complex systems can be a challenge. Currently, optimizations of σ and MM link partial charges for amino acids compatible with the CHARMM and Amber FFs are underway for ease of use.

11.7. Flexible Inner Region Ensemble Separator (FIRES). A QM/MM methodology in which a solute and the nearest water molecules are represented at a high *ab initio* level, offers a powerful strategy to study the hydration structure around small ions in the aqueous phase. However, one challenge with solvent molecules in hybrid QM/MM simulation is that they are free to diffuse away from the region of interest, and be replaced by MM solvent molecules that provide presumlably a less accurate model. To resolve this issue, FIRES was designed in which the ion and a fixed number of nearest water molecules form a dynamical and flexible inner region that is represented with a high level *ab initio* QM method, while the water molecules in the surrounding bulk form an outer region that is represented by a classical FF. Simulations with FIRES yield rigorously correct thermodynamic averages as long as the solvent molecules in the flexible inner and outer regions are not allowed to exchange. The method was used to study hydration structure around Na^+ and K^+,[698] and Mg^{2+} and Zn^{2+}.[715] To obtain a more efficient dynamical propagation algorithm, it is necessary to manage the computational cost of the QM part. To this end, a MTS dual-Hamiltonian propagation algorithm was designed by which the trajectory is propagated at every time step via a computationally inexpensive QM Hamiltonian, and then corrected less frequently using a more accurate and computationally expensive QM Hamiltonian.[716]

11.8. Combining QM/MM with Gaussian Process Machine Learning Potentials. The pyCHARMM[9] interface in CHARMM has facilitated advanced uses of QM/MM potentials in conjunction with Python-based ML potentials, including those described by neural networks[54] and Gaussian process regression (GPR).[717,718] Built upon multivariate Gaussian distribution of latent functions, GPR is a non-parametric, kernel-based stochastic inference ML approach that maximizes the likelihood of training data observation.[719] In simulations, GPR has been employed to model the relationship between molecular descriptors and the PES (reviewed in ref 363). Recently, GPR has been utilized to develop delta-ML potentials to improve SE-QM/MM free energy simulations.[717,718] By combining the AM1/MM potential in CHARMM and energy-based streaming sparse GPR (SSGPR) models, AI-QM/MM quality PES information can be learned along the string free energy paths in a data-efficient manner.[718] Using the extended-kernel GPR with derivative observations (GPRwDO), both energy and force matching can be employed to improve SE-QM/MM free energy simulations.[717] Figure 19A shows the PMFs for the Menshutkin reaction simulated at the AM1/MM level, before and after deploying the GPR correction model where the latter significantly alleviates overestimation of the free energy barrier

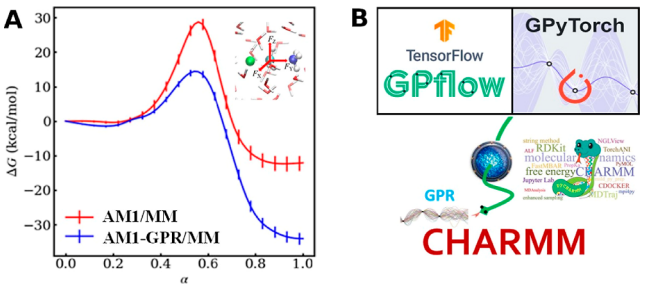


Figure 19. (A) PMFs of the Menshutkin reaction ($NH_3 + CH_3Cl \rightarrow NH_3CH_3^+ + Cl^-$) simulated at the AM1/MM and AM1-GPR/MM levels.[717] (B) The scheme of combining GPR Python libraries and CHARMM through pyCHARMM.

and the product free energy. In these QM-GPR/MM studies, GPR models trained using Python libraries such as GPflow[720] and GPyTorch[721] are deployed on the fly during MD simulations (Figure 19B). A Colab-based tutorial is available to demonstrate the use of the related Python libraries to train basic ML models for reactive systems.[722] A similar tutorial for training GPR models for QM/MM systems using CHARMM and pyCHARMM is currently under development.

11.9. Multistate Empirical Valence Bond (MS-EVB). The MS-EVB module of CHARMM[723,724] is an efficient method for representing reactive PES, e.g., in enzymes where a system moves from a reactant to a product topology (see ref 725 for implementation details). The most common approach to EVB involves a pseudo-Hamiltonian matrix $\mathbf{H}(\mathbf{q})$ constructed from two diabatic reactant (R) and product (P) basis functions, yielding a 2×2 matrix

$$\mathbf{H}(\mathbf{q}) = \begin{bmatrix} V_R + \epsilon_R & H_{1,2} \\ H_{1,2} & V_P + \epsilon_P \end{bmatrix} \tag{61}$$

where V_R and V_P are the potential energies of the reactant and product diabatic states at a given geometry $\mathbf{q}$, obtained from standard FF. ϵ_R and ϵ_P are constant diagonal energy shifts usually chosen to reproduce the known exo- or endothermicity of the reaction. The off-diagonal element $H_{1,2}$ couples the reactant and product basis functions, which is usually a simple function of atomic coordinates. In CHARMM, one can choose constants or 1D/2D Gaussians, which are functions of one or two distances between atoms. $\mathbf{H}$ can be diagonalized into $\mathbf{D} = \mathbf{U}^T\mathbf{H}\mathbf{U}$, where the diagonal matrix $\mathbf{D}$ contains the eigenvalues and $\mathbf{U}$ consists of eigenvectors of $\mathbf{H}$. Applying the Hellman-Feynman relation gives a matrix of Cartesian atomic forces

$$\mathbf{F} = -\frac{\mathrm{d}\mathbf{D}}{\mathrm{d}\mathbf{q}} = -\mathbf{U}^T \frac{\mathrm{d}\mathbf{H}}{\mathrm{d}\mathbf{q}}\mathbf{U} \tag{62}$$

which contains the gradient vector $\mathbf{F}_i$ for each adiabatic state corresponding to the i-th eigenvalue of $\mathbf{D}$ in increasing order of energy. $\mathbf{F}_0$ contains forces corresponding to the lowest eigenvalue λ_0, and is used for dynamics propagation on the adiabatic ground state.

The CHARMM-EVB implementation utilizes MPI to parallelize the energy and force calculation for each topological replica at any given time step, achieving near-linear scaling with the number of topological replicas so that the number of topological replicas is limited only by the number of MPI threads running on the given hardware.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

11.10. Reactive MD. Following chemical reactions in time and space is a central aspect of chemistry. For computer-based methods, *ab initio* MD methods at correlated levels are usually too prohibitive, in particular if statistically significant numbers of trajectories need to be run. Earlier and previous empirical efforts to describe bond breaking and formation include approaches based on bond order and bond strength.[726−729] Alternatively, chemical reactivity can be modeled as a linear combination of empirical energy functions describing two or multiple atom connectivities (reactant and one or several products) and to mix these representations. This leads to reactive PESs as in multistate adiabatic reactive MD (MS-ARMD).[730−732] Here, the PESs are mixed according to

$$V_{\text{MS-ARMD}}(\mathbf{x}) = \sum_{i=1}^{n} w_i(\mathbf{x}) V_i(\mathbf{x}) \tag{63}$$

The weights $w_i(\mathbf{x})$ are obtained by normalizing the Boltzmann distributed raw weights $w_{i,0}(\mathbf{x})$

$$w_i(\mathbf{x}) = \frac{w_{i,0}(\mathbf{x})}{\sum_{j=1}^{n} w_{j,0}(\mathbf{x})}, \qquad w_{i,0}(\mathbf{x}) = \mathrm{e}^{-(1/\Delta V)[V_i(\mathbf{x}) - V_{\min}(\mathbf{x})]} \tag{64}$$

where $V_{\min}(\mathbf{x})$ is the minimal energy for a given configuration $\mathbf{x}$ and ΔV is a characteristic energy scale (switching parameter). Per construction (*cf.*, Eq. 64), only surfaces within a few times of ΔV from $V_{\min}(\mathbf{x})$ will contribute to instantaneous configuration $\mathbf{x}$. ARMD mixes different PESs V_i by using Gaussian and polynomial functions around the crossing points between states by fitting to reference data such as the MEP.[732] Because the mixed PES $V_{\text{MS-ARMD}}(\mathbf{x})$ depends on energies of different states through weights w_i which in turn are analytical functions of the coordinates $\mathbf{x}$, energy-conserving MS simulations can be run using MS-ARMD.[732]

A more recent extension combines MS-ARMD[733] with VALBOND, a FF that allows to describe the geometries and dynamics of metal complexes.[734−736] The formulation is reminiscent of empirical valence bond theory[600] where diagonal terms are VALBOND descriptions of the states involved and off-diagonal terms describe the orbital overlap. MS-ARMD can also be combined with MM with proton transfer (MMPT),[737] to follow proton transfer in gas and condensed phases.[738−741]

In the gas phase, MS-ARMD was used to study reactions such as hydrogen transfer in the photodissociation of $H_2SO_4 \rightarrow H_2O + SO_3$[742] and other atmospherically relevant molecules by following vibrational excitation of the OH stretch,[743,744] the Claisen rearrangement reaction,[745] or to investigate Diels−Alder reactions.[746] Such studies provide insights into reaction mechanisms and relevant coordinates driving the process. As an example, for the Diels−Alder reaction between 2,3-dibromo-1,3-butadiene and maleic anhydride MS-ARMD emphasized the importance of rotations of the two reactants to reach the transition state.[746]

More recently, the unimolecular dissociation of vibrationally excited *syn*-CH_3CHOO to form OH and CH_2CHO was investigated (Figure 20).[382,389] For the reactant and product states, MS-ARMD performs close to the chemical accuracy (∼1 kcal/mol) whereas the MEP is described considerably more accurately (inset in Figure 20B). Atomistic simulations using the MS-ARMD PES are about 2 orders of magnitude more efficient than using a neural network-based PES and about 6 orders of magnitude faster than *ab initio* MD simulations at the MP2 level of theory at which the MS-ARMD PES was developed. In other words, MS-ARMD simulations can be run routinely with high quality and in statistically significant numbers, as exemplified in Figure 20C.

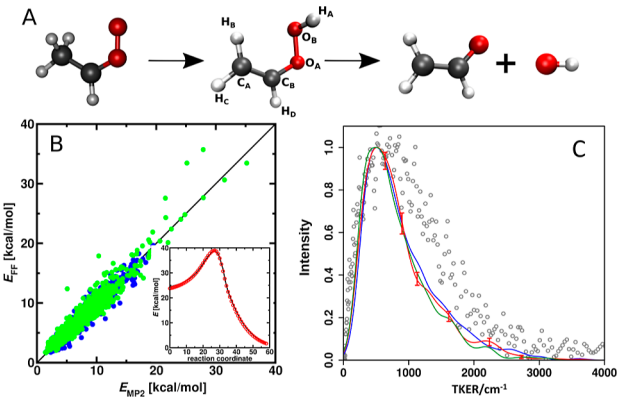


Figure 20. (A) OH-elimination following vibrational excitation of *syn*-CH_3CHOO. (B) Performance of MS-ARMD. The RMSEs between reference *ab initio* energies and the fitted FF for reactant (blue) and product (green) are 1.1 and 1.2 kcal/mol, respectively. Inset: MEP calculated with MS-ARMD (red circles) compared with reference calculations (black line). (C) Distribution of the total kinetic energy release from several thousand trajectories following CH-excitation with ∼2 quanta using the MS-ARMD PES with OO scission energies of 22, 25, and 27 kcal/mol (blue, red, green). Open symbols are experimental results.[387] Panels A and B reproduced from ref 382. Copyright [2021] American Chemical Society.

Finally, biological systems were also studied using a combination of RKHS-based PESs and empirical FFs,[747,748] which allowed structural interpretation of metastable states in MbNO and a molecularly refined understanding of ligand exchange (NO vs O_2) at the heme-iron in truncated hemoglobin.

11.11. Indirect QM/MM Free Energy Simulations. The alchemical free energy functionality, specifically the PERT module of CHARMM, was described in detail in the 2009 paper.[3] In addition to discrete intermediate states as a function of the coupling parameter λ, PERT supports slow-growth TI (SGTI) that changes λ incrementally at each step of the MD simulation. SGTI suffers from the Hamiltonian lag problem,[749] hence it is rarely used directly. A free energy difference obtained from SGTI should be treated as non-equilibrium work (NEW),[750] and the equilibrium free energy difference can be obtained by applying the Jarzynski equality[751] or Crooks theorem[752] to a sufficient number of SGTI runs.[753] Such calculations can be automated by CHARMM's scripting language, as illustrated below.

PERT fully supports CHARMM's multiscale capabilities (MSCALE module).[262] This makes it possible to compute free energy differences between two descriptions of a system, such as an MM description on one hand and a hybrid QM/MM description on the other hand. Most standard applications of alchemical free energy simulations, such as the calculation of relative binding free energies, employ equilibrium methods (TI,[329] Bennett's acceptance ratio method, BAR,[754] or its multistate extension MBAR[12]). NEW based methods are also used.[169,755] For such traditional applications, it is unclear whether equilibrium or non-equilibrium techniques are more efficient. The situation is different when one has to compute free energy differences between levels of theory, as is the case

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

for the so-called indirect cycle QM/MM alchemical free energy simulations.

The calculation of free energies when using a QM/MM description poses two challenges: (1) It is slow, making it difficult to achieve sufficient sampling; and (2) standard recipes to realize alchemical transformations, such as soft-core potentials do not work with QM/MM Hamiltonians.[297,377,756] Indirect cycles can circumvent both issues. The basic idea is illustrated using the calculation of an absolute solvation free energy (Figure 21). Instead of computing $\Delta G_{\mathrm{solv}}(\mathrm{QM/MM})$

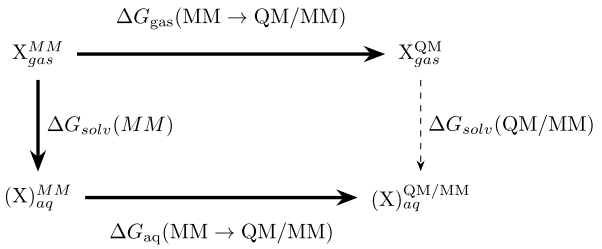


Figure 21. Illustration of an indirect cycle to compute a free energy difference at the QM/MM level of theory.

directly, one computes $\Delta G_{\mathrm{solv}}(\mathrm{MM})$ where soft-core potentials, *etc.*, can be used without restrictions. In addition, free energy differences $\Delta G_{\alpha}(\mathrm{MM} \rightarrow \mathrm{QM/MM})$ (α: gas or aq) both in the gas and aqueous can be calculated, which yields

$$\begin{aligned}\Delta G_{\mathrm{solv}}(\mathrm{QM/MM}) &= -\Delta G_{\mathrm{gas}}(\mathrm{MM} \rightarrow \mathrm{QM/MM}) + \Delta G_{\mathrm{solv}}(\mathrm{MM}) \\ &\quad + \Delta G_{\mathrm{aq}}(\mathrm{MM} \rightarrow \mathrm{QM/MM})\end{aligned} \tag{65}$$

The calculation of $\Delta G_{\alpha}(\mathrm{MM} \rightarrow \mathrm{QM/MM})$ is challenging,[757] but NEW-based approaches have been shown to be reliable and efficient.[758−761]

The combination of PERT and MSCALE enables the calculation of NEW values for transitioning from, e.g., a MM to a QM/MM description using solely CHARMM's scripting language. By inserting as few as two hundred of such work values obtained from trivially parallel simulations, into the Jarzynski equality, $\Delta G_{\alpha}(\mathrm{MM} \rightarrow \mathrm{QM/MM})$ can be calculated accurately and efficiently in most cases.[760,761] The NEW switches require equilibrium configurations sampled in the canonical ensemble, i.e., at the low level of theory. Restart files are saved at regular intervals, from which 2−5-ps long independent NEW switching simulations start in parallel.

A self-contained example illustrating this procedure is available at Zenodo.[762] While the example uses the SCC-DFTB method as the high level theory,[163] changes required for, e.g., a true DFT method are trivial. The key step is using MSCALE to employ one master (control) job and two slave jobs. The latter are responsible for computing energies and forces at the respective MM and QM/MM levels. The master process is primarily responsible for mixing forces/interactions and integrating the equations of motion. Care is needed to avoid double-counting if additional restraints or similar terms are used. The switching itself is realized as a SGTI calculation of the PERT module. By default, one switches linearly from λ = 0 to λ = 1 in 1,000−5,000 MD steps. The final result of each switch is the NEW value W, which can be saved or extracted automatically within the CHARMM script. Full automation can be achieved by calling the relevant CHARMM jobs from e.g., the Unix shell, a Python script, *etc.*

If the convergence of results obtained by the Jarzynski equality[751] is in doubt, one can also use the Crooks theorem.[752] In this case, a QM/MM-level equilibrium simulation is needed, followed by switches in the QM/MM → MM direction. The latter are again trivially parallel. While the computational cost of generating the initial configurations from the equilibrium QM/MM simulation is high, this workflow is still significantly more efficient than equilibrium-based approaches, which would require adequate sampling at each intermediate state (typically ten or more).

12. BOUNDARY CONDITION, SYSTEM PREPARATION, AND TRAJECTORY ANALYSIS

In addition to performing simulation itself, an ability to prepare the simulation system in a desired initial state, impose appropriate boundary conditions or constraints, and analyze simulation trajectories are essential for making scientific discoveries. CHARMM has an extensive set of tools available for this purpose. Presently, at least 25% of more than 1.17 M lines of the CHARMM source code belong to this category. Example applications of the recently developed methods described below are given in references therein. A vast array of other existing tools can be found in the documentation as well as the example 'Testcase' input scripts provided in the CHARMM package.

12.1. Simulation and Analysis of Membrane Proteins. By virtue of its considerable functional flexibility, CHARMM has been a tool of choice in many studies of ion channels and membrane proteins. A theory was developed and implemented to account for the membrane potential and its representation by a constant electric field in computer simulations.[763] It was subsequently used in studies of the Kv1.2 potassium channel[764] and the voltage sensing domain of the voltage-sensitive phosphatase from *Ciona intestinalis*.[765] Ion permeation through various channels was characterized,[766−769] and the fundamental principles governing ion selectivity were explored.[770,771] Computational methods with empirical energy restraints were developed to exploit information from low-resolution experimental data in structural refinement of membrane proteins. A particular attention was given to electron paramagnetic resonance (EPR) accessibility data,[772−774] and double electron−electron resonance (DEER) technique that reports distance distribution between spin labels.[775−777] The EPR/DEER methodology is also supported by CHARMM-GUI for easy setup of restrained simulations.[95] Using these methods, the structure of various ion channels and membrane transporters were refined on the basis of EPR experimental data.[772,773,778−780] Energy restraints were also developed to exploit information from mutational cross-link data,[781] which resolved ambiguities about the conformation of the resting state of the voltage sensing domain of potassium channels.[782,783]

12.2. Coordinate Unwrapping for Diffusion Constants. In CHARMM, fractional coordinates are used to unwrap trajectories, which yields the same diffusion constants as the more recent "toroidal view preserving" method.[784] It also highlights the need to correct calculated diffusion constants for PBC artifacts, especially those in lipid bilayers. The translational diffusion constant D is typically calculated using the Einstein relation

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

$$2nD = \lim_{t\to\infty} \frac{\mathrm{MSD}}{t} \tag{66}$$

where n is the spatial dimension, MSD is the mean-squared displacement, and t is simulation time.[785] In simulations with finite periodic box, it is standard to "wrap" or "image" positions of molecules such that a molecule crossing a unit cell boundary is translated to the opposite side. This effect must be removed via "unwrapping" for MSD calculation,[786] which is done in either Cartesian or fractional coordinates.

Constant volume (NVT) or constant energy (NVE) ensembles are recommended for diffusion calculations, as they minimize perturbations to dynamical variables. In MD codes tracking atomic virials like CHARMM, another advantage of using NVT is that the system viscosity and particle diffusion constants can be computed from the same trajectory to compare with experiment.[788] Unwrapping is straightforward under constant volume, where Cartesian coordinates can be used. However, it can be confounded by volume fluctuations in NPT simulations. For example, the heuristic unwrapping scheme (in which the position of a particle is unwrapped by comparing its current wrapped position to its unwrapped position at the previous time step) used by several MD software packages with Cartesian coordinates was shown to introduce cumulative errors in molecule's position and calculated MSDs.[789] A new method called the "toroidal view preserving scheme" was proposed by the same group to correctly unwrap such simulations.[784]

CHARMM avoids the preceding problem by always unwrapping coordinates in fractional space (*i.e.*, the space where each unit cell vector is transformed into 3 orthonormal vectors and positions are mapped onto this lattice) before projecting the coordinates back in Cartesian space. In fractional coordinates, the box fluctuations are removed, and MSD vs t plots with the correct slope are produced. Noise at longer simulation times can be mitigated by projecting the coordinates back into Cartesian space using the average unit cell vectors; the keyword for this operation in CHARMM is XFLUC.

The MSD is often computed using multiple time origins as a difference correlation function, where any deviation is indicative of a problem.[784] For comparison, Figure 22 shows the single time origin MSD vs t for 1340 TIP3P waters at 20 °C for NPT and NVT simulations. The NPT simulation was unwrapped in four ways: Cartesian coordinates (non-CHARMM), CHARMM fractional coordinates, CHARMM fractional coordinates with average box dimensions (XFLUC), and toroidal view preserving.[784] The NPT simulation unwrapped with the Cartesian scheme (violet) shows a significant accumulation of error and deviation from linear behavior after 500 ns. The increase in noise with simulation time for fractional space unwrapping (black) is essentially eliminated by projecting back onto the average unit cell vectors (blue). The resulting MSD vs t is practically indistinguishable from the results obtained with NVT (orange) or the toroidal view preserving scheme (green). Diffusion constants obtained from any of the non-Cartesian methods are therefore statistically equal.

Note that even for Cartesian unwrapping, the first several hundred nanoseconds appear unaffected. Comparison of the wrapping frequency distribution for water and for self-diffusion in a DPPC bilayer shows stark differences (Figure 23). Wrapping events for water are 2 or 3 orders of magnitude more

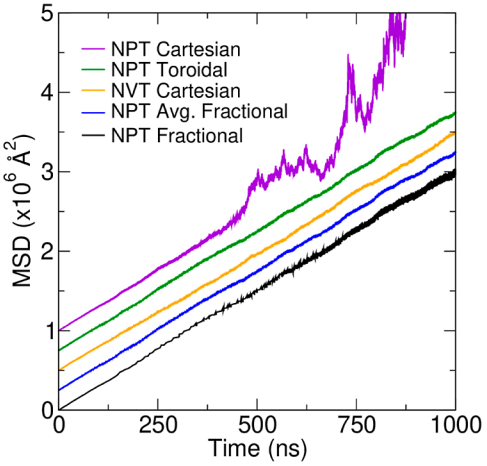


Figure 22. MSD vs t for 1-μs NVT and NPT simulations of 1340 TIP3P waters. The plots are offset by intervals of 0.25×10^6 Å^2 on the y-axis to better distinguish them. From top to bottom, violet: NPT with Cartesian unwrapping; green: NPT with toroidal view preserving scheme; orange: NVT with Cartesian unwrapping; blue: NPT with unwrapping in fractional space and using average unit cell vectors (CHARMM XFLUC); black: NPT with unwrapping in fractional space (CHARMM method). Trajectories were run with OpenMM and analyzed with CPPTRAJ version 6.19.3.[787]

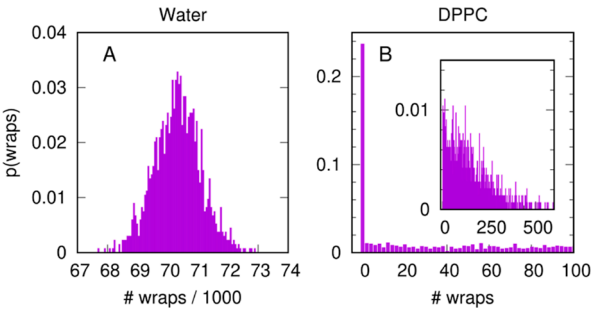


Figure 23. Probabilities of wrap counts for water molecules from a 1-μs NPT MD simulation of TIP3P water and for a bilayer containing 288 DPPC molecules from 5 replicate 400-ns MD simulations.[571] The inset omits the large peak for DPPC with zero wraps; the mean number of lipid wrap counts is 111.5.

frequent than lipids, and the distribution is approximately Gaussian. The distributions for lipids are Poisson-like with zero being the most frequent value.

Errors obtained from the Cartesian unwrapping scheme are related to the number of wrapping events for a given molecule. Evaluation of lipid diffusion in bilayers from published simulations[571] shows that for larger, slower moving molecules the unwrapping method had very little effect on the results. Table 3 shows that the standard deviation over five replicate simulations (bottom row) is an order of magnitude larger than that over the three different unwrapping methods (last column).

Diffusion constants are also affected by the PBC that causes underestimates of the infinite system-size value. PBC errors in diffusion constants in isotropic systems are relatively modest, approximately 10%, and can be corrected by the Yeh–Hummer formula.[790] The PBC correction for diffusion in lipid bilayers requires the periodic Saffman–Delbrück model.[791,792] It is quite large for lipid self-diffusion, approximately 3-fold for

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 3. D_{MSD} (in Units of 10^{-7} cm^2/s) for NPT DPPC Bilayers with Drude2023 FF, with Cartesian (Cart.), Fractional (Frac.), and Toroidal (Tor.) Unwrapping

Rep #	Cart.	Frac.	Tor.	Avg	Std
1	0.71	0.71	0.70	0.71	0.001
2	0.73	0.73	0.73	0.73	0.001
3	0.74	0.74	0.75	0.74	0.001
4	0.67	0.67	0.67	0.67	0.003
5	0.69	0.69	0.69	0.69	0.003
Avg	0.71	0.71	0.71	0.71	0.001
Std	0.03	0.03	0.03		

bilayers with 288 lipids, and should not be overlooked. Our overall recommendations for unwrapping are the following:

- The Cartesian-based method should not be used when the box dimensions can change during a simulation. Though for relatively short simulations of slowly diffusing particles the errors are not substantial (Table 3), accurate methods are readily available and should be used.
- The XFLUC method in CHARMM should only be used when the box dimension fluctuations are less than approximately 15%. This is its major limitation. Though XFLUC would still yield smooth plots, the slope (and hence D) could be incorrect if the aspect ratio changes significantly.
- Although the toroidal view preserving method provides a smoother result than with CHARMM fractional coordinates, both methods should give correct answers if the average slope is used.

12.3. Calculation of Pressure Profiles in Lipid Bilayers. The lateral pressure profile provides a detailed view of the forces within a planar lipid bilayer with respect to the average bilayer normal (typically z-axis). CHARMM provides capability for estimating the spontaneous curvatures in symmetric[557] and asymmetric bilayers[44] and the difference in leaflet surface tensions (or differential stress) in asymmetric bilayers.[44,793] For a simulation system Ω, the volume-averaged virial stress tensor, $\boldsymbol{\sigma}$, is[794]

$$\boldsymbol{\sigma} = -\frac{1}{V}\left(\sum_{i\in\Omega} m_i\vec{v}_i \otimes \vec{v}_i + \sum_{\{i<j\}\in\Omega} \vec{F}_{ij} \otimes \vec{r}_{ij}\right) \tag{67}$$

where V is the volume of Ω and m_i and $\vec{v}_i$ are the mass and velocity of atom i. The symbol $\otimes$ represents the tensor product between two vectors, $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$, and $\vec{F}_{ij}$ is the force exerted from atom j on atom i. The first and second terms on the right-hand side of Eq. 67 are the kinetic and configurational contributions, respectively. In the configurational virial stress, the contributions from periodic images must be considered for non-bonded interactions. The pressure tensor ($\mathbf{p}$) is defined as the negative of the stress tensor, i.e., $\mathbf{p} = -\boldsymbol{\sigma}$. For an isotropic system, the bulk pressure is $P = -\mathrm{Tr}(\boldsymbol{\sigma})/3$. CHARMM calculates the virial stress (and thus the pressure) using Eq. 67. While constant pressure can be emulated using Monte Carlo barostat,[34] the virial is needed to calculate transport properties such as viscosity.[788]

Derivatives of the free energy with respect to virtual transformations[795] of the periodic box can be computed with the lateral pressure profile $p_L(z)$ typically along the z-direction, normal to the membrane surface

$$p_L(z) = p_T(z) - p_N(z) \tag{68}$$

where $p_T(z)$ and $p_N(z)$ are the tangential and normal components of the pressure tensor, respectively. The zeroth moment of the profile is the tension (the derivative of the free energy with respect to area) while the first moment is the derivative of the free energy with respect to the curvature. Interpreted through the Helfrich/Canham Hamiltonian,[796,797] it provides a convenient route to calculate spontaneous curvature of the leaflet.[798]

For planar lipid bilayers, the slab geometry is a convenient choice for lateral pressure profile calculations

$$p_T = (p^{xx} + p^{yy})/2, \quad p_N = p^{zz} \tag{69}$$

Here, p^{xx}, p^{yy}, and p^{zz} are the diagonal elements of $\mathbf{p}(z)$. Since a typical lipid bilayer cannot support in-plane shear strain, there is no off-diagonal coupling of x and y.

Denoting the z-dimension of the simulation box as L_z, the bilayer and leaflet surface tensions are given as zeroth-moments of $p_L(z)$:

$$\gamma = -\int_{-L_z/2}^{+L_z/2} \mathrm{d}z p_L(z),$$
$$\gamma_1 = -\int_{0}^{+L_z/2} \mathrm{d}z p_L(z),\ \gamma_2 = -\int_{-L_z/2}^{0} \mathrm{d}z p_L(z) \tag{70}$$

Without any external force, the bilayer tension must vanish ($\gamma = 0$). While leaflets are tensionless ($\gamma_1 = \gamma_2 = 0$) in symmetric bilayers, leaflet tensions and their difference ($\Delta = \gamma_1 - \gamma_2$) in asymmetric bilayers do not necessarily vanish.[799]

The leaflet spontaneous curvature c_0 of a (planar) symmetric bilayer can be calculated from the first moment of $p_L(z)$

$$k_c c_0 = \int_{0}^{L_z/2} \mathrm{d}z z p_L(z) \tag{71}$$

where k_c is the bending modulus of the leaflet. Spontaneous curvatures of asymmetric bilayers can be calculated with a generalization of Eq. 71.[44,800]

While the bulk pressure tensor can be readily calculated using Eq. 67, the calculation of the local pressure tensor (including the lateral pressure profile) is complicated by the need to assign each contribution to the virial locally in space (here, z). Briefly, a vector *contour* is integrated from one force center i to the other center j, creating a 3D function whose gradient is a Dirac-δ function at each end point multiplied by the force. With this requirement met, the pressure tensor contains spatial correlations of force such that the work required to reshape the box by a virtual deformation can be computed. Many choices of contour satisfy this requirement, leading to the inherent ambiguity of the profile unless complemented by a virtual deformation yielding an observable that resolves the ambiguity.[795] Because of this complexity, the pressure profile calculation has not been supported in most other simulation programs.

The LOPR module in CHARMM calculates p_T in Eq. 69 where their profiles from the full electrostatics can be obtained by either the Ewald sum[801] or by the PME method.[176] Presently there are only two programs in addition to CHARMM that support the pressure profile from full electrostatics: NAMD (by Ewald sum in post analyses) and GROMACS (by PME[802] in a branch version). Due to the limitation of the Harasima contour[801] employed for the virial

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

calculation, the normal component p_N (Eq. 69) is not calculated but can be evaluated in post analysis. Furthermore, a planar bilayer cannot support heterogeneous normal pressure. For a bilayer without any external force ($\gamma = 0$), p_N is calculated from Eqs. 68 and 70 as

$$p_N = \frac{1}{L_z}\int \mathrm{d}z p_T(z) \tag{72}$$

The pressure profiles $p^{xx}(z)$ and $p^{yy}(z)$ are typically calculated by binning the z-dimension.[803] For accurate profiles, the bin size is set typically to ~1 Å. Alternatively, p^{xx} and p^{yy} can be calculated in a binless manner using Fourier series

$$p^{\alpha\alpha}(z') = \frac{w_{00}}{2} + \sum_{n=1}^{N}\sum_{l=0}^{1} w_{nl}\cos(2\pi n z' - \phi_l)$$

$$w_{00} \equiv 2\int_{-1/2}^{+1/2} \mathrm{d}z' p^{\alpha\alpha}(z')$$

$$w_{nl} \equiv 2\int_{-1/2}^{+1/2} \mathrm{d}z' p^{\alpha\alpha}(z')\cos(2\pi n z' - \phi_l) \tag{73}$$

where α represents x or y and $z' = (z - z_{cm})/L_z$ is the fractional z-coordinate with respect to the bilayer center z_{cm}. N is the order of the Fourier series, w_{nl} are Fourier coefficients, and $\phi_0 = 0$ and $\phi_1 = \pi/2$ are phase shifts for even and odd series. The LOPR module supports both methods for the lateral pressure profile calculation. The Fourier series method is currently supported only in CHARMM, where accurate pressure profiles can be obtained with a moderate number of coefficients, $N \sim 20$ for typical bilayers with $L_z \sim 80$ Å. For simulation systems with larger L_z, larger N is required.

The LOPR module supports both on-the-fly and post analysis calculation of $p_T(z)$ including the full electrostatics via the PME method. This allows efficient resampling where sparsely sampled coordinate and velocity trajectories from various programs including CHARMM, NAMD, and OpenMM (with a customized Velocity Reporter for trajectory generation)[793] can be utilized (Figure 24A,B). In the resampling approach, multiple short CHARMM simulations for chosen frames can be run simultaneously, which can be easily realized in typical computational resources. If there are sufficient samples from long simulation times, one can also calculate the lateral pressure profile in a post analysis using a single-step dynamics for each frame with an integration time step shorter than the one used for the original simulations. The post analysis method is faster than the resampling method, and yields sufficiently accurate results from 500-ns trajectories saved at every 10 ps (Figure 24B). For an asymmetric bilayer, the pressure profile is also asymmetric (Figure 24C) from which the non-vanishing leaflet tensions are calculated from Eq. 70.

The LOPR module currently does not support LJ-PME[564] (Section 8.5) and the polarizable Drude model[571] which will be supported in future updates. Additionally, it is implemented only for CPU calculations without DOMDEC[25] (Section 2.2), which results in poor scalability over multiple nodes. Thus, the current LOPR module is suitable practically for only single node-jobs, and parallelization of the code will greatly improve its performance.

12.4. $P2_1$ Periodic Boundary Condition. A novel approach utilizing the $P2_1$ PBC has been implemented to relax differential stress between the leaflets during MD simulations of lipid bilayers. The inherent difficulty in accurately estimating the number of lipids in each layer *a priori* gives rise to the differential stress. The widely used P1 PBC involves tessellating the simulation space with translated images of the box. As an atom exits the simulation box, it is replaced by its image located on the opposite face. Unlike P1, the $P2_1$ PBC introduces a half-screw symmetry between the images. In this scheme, the image of the simulation box is not a mere translated copy but a 180°-rotated image translated along the same screw axis. In the CHARMM non-DOMDEC version (DOMDEC is explained in Section 2.2), the screw axis could be oriented in any direction. However, with the Extended Eighth Shell (EES) method[43] in DOMDEC, the screw axis is constrained to the x-axis for enhanced performance. The $P2_1$ symmetry operation is denoted as $(x + 1/2, -y, -z)$, representing a half-unit cell length translation along the x-axis and reflection along the y- and z-axes. Reflection along two perpendicular axes is equivalent to 180° rotation along the

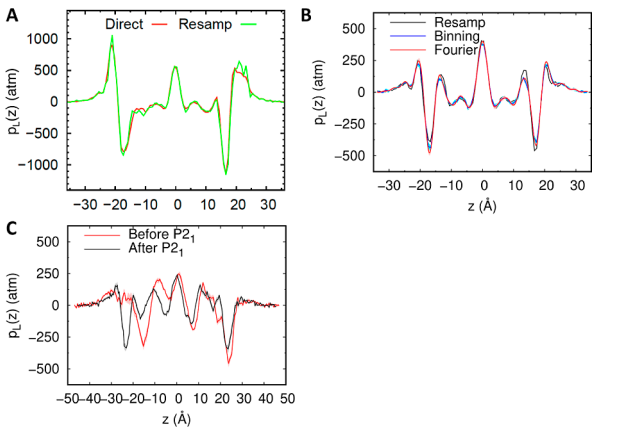


Figure 24. (A) Comparison of pressure profiles from a 100-ns CHARMM simulation of palmitoylsphingomyelin bilayer with complete sampling (Direct) or via the 10% resampling (Resamp) of 100-ps intervals spaced 1-ns apart, using restart files from the fully sampled simulation. The values of the first moment and their standard errors are 0.190 ± 0.009 and 0.192 ± 0.004, respectively. (B) Pressure profiles of a bilayer composed of 72 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine at 1 atm and 323 K. Data from resampling were obtained by the binning method with 100 bins from the previous CHARMM simulations (black), where 100-ps resampling was done for every 1 ns.[558] Post analysis using single-step dynamics was tested using 500-ns OpenMM NPT trajectories of the same bilayer, where the pressure profiles were calculated for each frame by both the binning (blue) and Fourier series (red) methods. The numbers of bins and Fourier coefficients were set to 80 and $N = 20$ (Eq. 73), respectively. Standard errors from five 100-ns blocks are shown in cyan and pink areas, respectively, which are smaller than the line thickness except near dips. (C) Pressure profiles for a 1,2-dinervonoyl-*sn*-glycero-3-phosphocholine bilayer with a model peptide of 9 monomeric units of gramicidin A in the upper leaflet at a peptide area fraction $\phi \sim 0.40$ before (red) and after $P2_1$ equilibration (black). The $P2_1$ PBC in CHARMM allows lipid translocation between bilayer leaflets, which reduces area stress and alters the lateral pressure profile (see Section 12.4). The bilayer midplane was set to $z = 0$. Pressure profiles from the last 300-ns trajectories of 5 replicas from OpenMM were averaged for each asymmetric bilayer, whose standard errors are shown as pink and gray areas (typically smaller than the line thickness). Pressure profiles were calculated by single-step dynamics for each frame of trajectories with 200 bins along the z-direction. Panel A reproduced from ref 804. Copyright [2014] Cell Press. Panel C reproduced from ref 44. Copyright [2023] Wiley.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

screw axis where lipids departing from the top layer along the *x*-axis (*yz*-faces) reenter the cell in the bottom layer, and *vice versa*. However, since only the *x*-axis is allowed as the screw axis, lipids leaving the cell along the *xz*-faces reenter along the same leaflet. This can be visualized as a torus-shaped structure along the screw axis where the top layer transitions to the bottom layer in the neighboring image cell. A bilayer simulation starting with 108 lipids on the top layer and 92 lipids in the bottom equilibrates to 100 lipids in both layers using this method.[43]

The scaling performance of the EES method for $P2_1$ simulations is similar to that for DOMDEC in P1 simulations. Compared to the latter, there is slightly larger import volume during the message transfer among nodes during the direct space calculations. However, by restricting the screw axis to the *x*-axis, there is a minimal impact of the larger message size on the overall performance. The reciprocal space calculations are done by distributing the charge on the full unit cell. However, forces are calculated only in the asymmetric unit, and extra bookkeeping is also performed to rotate the forces and velocities as the images are 180°-rotated along the *yz* faces.

While it may first appear that the exchange of lipids between the layers would allow only symmetric bilayer simulation, the $P2_1$ PBC is specifically useful for setting up asymmetric bilayers by restraining specific lipids to their original leaflet and redistributing others between the leaflets through $P2_1$ PBC. The differential stress or the difference in surface tension between the two layers is a consequence of the intrinsic bending and the asymmetric lipid packing. Methods for simulating asymmetric bilayers can be categorized as lipid-based, leaflet-based, and bilayer-based.[793] In the lipid-based approach termed APL, surface areas of the top and bottom are matched using the area per lipid from homogeneous lipid bilayers. This is the simplest and the most commonly used approach that disregards any coupling between the bilayers and assumes ideal mixing of lipids. The leaflet-based approach termed SA minimizes the differential area strain between the leaflets. It first equilibrates symmetric bilayers corresponding to each leaflet composition and then combines one leaflet from each bilayer. While this removes the differential strain, it also disregards coupling between the two leaflets. The bilayer-based approach termed 0-DS removes the differential stress by adjusting the number of lipids. When these approaches are followed by $P2_1$, agreement in mechanical properties significantly improves among APL/$P2_1$, SA/$P2_1$, and 0-DS/$P2_1$.[793] These findings align with a theoretical framework emphasizing the intricate interplay between bending and asymmetric lipid packing. Torque balance conditions and stress indices provide theoretical support, showcasing promising results for $P2_1$ simulations in capturing lipid asymmetry observed in biological membranes.

As another example of the importance of $P2_1$ for setting up bilayer simulations, the curvature induced by a peptide in the so-called "peptide-asymmetric bilayer" was studied.[44] In these simulations, while both layers contained the same types of lipids, the asymmetry was induced by the presence of the peptide in the cis-leaflet. A series of gramicidin A (gA)-based peptides were simulated: a single monomer, a fused tetramer, a fused nonamer, nine gA monomers, and a fused tetramer of a gA mutant whose Trp residues were replaced by Gln. These assemblies were used to investigate effects of the size of chemically similar peptides spanning a single leaflet. Utilizing the APL method mentioned above, systems were created at three distinct peptide area fractions. Subsequently, equilibration employing the $P2_1$ PBC was performed. The P1 simulation showed significant condensation of lipids in the trans leaflet, resulting in large differential stress and bilayer bending moment, which relaxed via the exchange of lipids between leaflets in $P2_1$ simulation (Figure 24C).

12.5. Primary Hydration Shell (PHS) Model. While the representation of solvent surrounding macromolecules should closely approximate physical reality in MD simulations, the common use of a sizable volume of solvent with PBC is computationally expensive. A PHS consisting of 2−3 layers of explicit water molecules around a protein may be sufficient to maintain the conformational stability and dynamics of macromolecules. The initial work[631] on the PHS model has been refined in two stages. First, the method was tested with hen egg lysozyme, where good agreement with the protein and solvent behavior was observed, including Lipari−Szabo order parameters for N−H main-chain and $N-H_2$ side-chain motions on the ps−ns time scale in simulations of 25−150 ns, compared to full PBC treatments.[805] The original PHS method has a modest half-harmonic restraint of waters to their nearest protein atom, should they become more distant than a threshold value (5.8 Å by default). As a part of this work a simpler GEO restraint has been implemented which saves computer time as it is calculated relative to three perpendicular principal axes that follow the protein frame. The GEO approach follows global conformational changes and is similarly good for the lysozyme tested.

The PHS method was subsequently refined to overcome issues when applied to larger systems.[806] A neighbor list was implemented to efficiently track the nearest protein atoms to the water oxygen atom, to avoid calculating distances at each step. Also, an asymmetric harmonic potential instead of a half-harmonic one was used to ensure correct water density close to the boundary. In addition, pressure control was implemented, and the confining potential was scaled to keep waters near hydrophobic residues. This approach showed a 14-fold reduction in computing time for a 82-kDa protein. Future developments of the PHS model should handle situations with extensive structural changes[807] and association between proteins where treatment of long-range force is important.[808]

12.6. Hydration Map. Surface hydration of proteins and nucleic acids are important for their biological function,[809] which has long been a subject of computer simulation.[810] For measuring location-dependent average behavior of water molecules near biomolecular surfaces, the COORdinates SMAP (Solvation MAP) command has been implemented in CHARMM.[811,812] It divides the simulation water box into a grid of cubic cells (default size: 0.7 $Å^3$, half the radius of a water molecule), and locally calculates time-averaged properties of water within each cell. To account for protein motion during simulation, coordinate frames are aligned to a reference structure so that the calculated map is relative to the surface. In the current implementation, local water density (SDENsity), translational diffusion coefficient (DIFTrans), and the average number of hydrogen bonds formed by water molecules (HBONd) can be calculated. By default, water oxygen atoms are used for calculation. Other atoms such as ions can be selected as 'solvent' atoms, to build the corresponding maps. This capability has been used to find the preferred location of sodium ions around a double-stranded DNA as the center of the minor groove, between the 'double water spines.'[812]

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The solvation map can be saved as a data file for further analysis, or an MRC format electron density map for visualization, e.g., by using the UCSF ChimeraX.[813] Since coordinate frames are oriented to a reference structure, calculated water densities around flexible loops may become low. To examine hydration around a moving loop, a separate solvation map can be built by selecting only the loop as the orientational reference.

The water density map can be used to calculate the solvation free energy.[814] For N cells surrounding the protein under consideration, if the water density of cell i during the simulation is ρ_i and denoting the bulk water density as ρ_b, the free energy of water for N cells is

$$G_{\text{solv}} = -k_B T \sum_{i}^{N} \ln(\rho_i/\rho_b) \tag{74}$$

Since the size of a cell is smaller than that of a water molecule (0.7 Å by default), a high water density at cell i means that the cell is visited more frequently rather than water molecules are packed more tightly. The underlying idea for Eq. 74 is that a more frequently visited cell has a lower free energy (a favorable location) compared to the bulk water. To use Eq. 74 in practice, cells corresponding to the first hydration shell are selected, for which a distance cutoff of 4.5 Å from heavy atoms of the protein, and water density cutoff of 0.034 $Å^{-3}$ (*cf.*, ρ_b = 0.0333 $Å^{-3}$) are used.[814] Compared to the popular grid inhomogeneous solvation theory (GIST),[815,816] the above method does not require the protein to be constrained during simulation, where constraint on proteins drastically alters surface hydration. And the measured solvation energy values are in physically more reasonable range compared to those from GIST.[814] Calculation of the density-based solvation free energy is being implemented in CHARMM as the COORdinates SMAP SLVE command.

12.7. Conformational Entropy. Conformational entropy is an essential component of the conformational free energy of a biomolecule. A widely used class of entropy calculation methods rely on quasi-harmonic approximation where frequencies of different vibrational modes are used to estimate conformational entropy.[817,818] However, they cannot account for transitions between states. With increases in computational power, more direct evaluation of entropy from distributions of degrees of freedom (DOF), in particular, backbone and side-side chain dihedral angles has become possible. If all N DOFs are treated independently, the cost of entropy calculation scales linearly with N. However, since DOFs can be mutually correlated (e.g., by correlated motion of side chains forming contacts), higher order corrections should be made, which amounts to calculating multivariate histograms. The maximal information spanning tree (MIST) approach systematically handles higher order corrections in such a way that the estimated entropy monotonically approaches its asymptotic value.[819,820] It should be noted that, an accurate calculation of higher order terms requires a greater number of coordinate frames since for n-th order correlation, the total number of bins for the histogram scales with N^n. With a limited number of coordinate frames, bins will be sparsely populated, leading to an increased statistical uncertainty. In practice, MIST calculation up to the second order is good for most purposes since it provides statistically reliable result, fast to calculate, and correlations beyond the second order do not usually contribute significantly.

By limiting DOFs to an amino acid side chain, MIST can be used to estimate the side-chain entropy of individual residues. In this case, since N is small (e.g., total number of dihedral angles of a side chain), calculation up to the third order MIST can be done for higher accuracy. This method has been used to calculate changes in the side-chain entropy for binding of proline-rich ligands to an SH3 domain.[821] In this study, "entropy hotspots" were identified where the side-chain entropy of remote residues in the SH3 domain increases upon ligand binding. This arises from the rearrangement of contacts across the protein's surface that makes the side chains of entropy hotspots to become more mobile upon ligand binding. While initially developed as a separate code,[821] the COORdinates MIST command is currently being implemented in CHARMM.

12.8. Identifying Non-Polar Contacts. In CHARMM, hydrogen bonds can be readily identified by the COORdinates HBONd command, which works either for a single structure or across coordinate frames. In comparison, identifying non-polar contacts has been less established. It is often desirable to determine non-polar contacts at the level of individual residues rather than between pairs of atoms. Measuring distances between C_α atoms or centers of mass between side chains of non-polar residues and applying an *ad hoc* cutoff distance does not provide an accurate picture of non-polar contacts. CHARMM now has the COORdinates DISTance RESIdue command. Without the RESIdue keyword, all pairwise distances between two groups of selected atoms are reported. With the RESIdue keyword, pairwise minimum distances between residues in the two groups (e.g., between two domains) are reported. To identify non-polar contacts, selecting atoms with the absolute value of charges less than $0.3e$ ($e = 1.6 \times 10^{-19}$ C) and distance cutoff of 3 Å can be used. These are based on charges of non-polar hydrogen atoms and their van der Waals radii (1.32 Å). In this way, physical contacts between non-polar residues can be identified and further processed to analyze their dynamics, i.e., occupancy, formation, and breakage.[822,823]

12.9. Vectorial Analysis of Long-Range Concerted Motions in MD Trajectories. Long-range concerted motions in proteins and other biomolecules is best captured with a correlation coefficient (CC) based on covariance using Euclidean distances between entries of a position-vector time series. This coefficient, DCOR, is a vector equivalent of Pearson's CC or a generalized CC.[824] The relative accuracy of DCOR is established by an assessment conducted using vector displacements generated with a known CC.[825] DCOR is least sensitive to angular variation between two vectors compared to Pearson's CC or a vector CC. Nor is DCOR as sensitive to large variations between vector components compared to the (scalar) generalized CC, which was found to give inflated CCs relative to the actual values when only one of the vector components is highly correlated.

The DCOR value between any two vector time series can be computed using the CORREL module. The vector dimensions need not be equal. For each time series, a matrix of intravector Euclidean distances between all pairs of time points in the series is used to calculate the covariance. DCOR reflects both linear and non-linear correlations,[825] and can detect long-distance concerted motions that neither Pearson's CCs nor the generalized CCs can reveal.[826]

12.10. Other Updated Preparation and Analysis Features. *12.10.1. System Generation.* The residue sequence

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

to be used in the generation of a segment in the PSF can be read from the ATOM and/or HETATM records in a PDB file with a specified chain or segment ID to the READ SEQU PDB command. If there are residues in the PBD file that do not exist in the RTF (Residue Topology File), or with names that differ from those used in the RTF, it is possible to either skip those residues or map the name in the PDB to the name in the RTF. The sequence can also be read from the SEQRES records, which is useful when there are missing residues in the PDB-file. These enhancements allow the generation of a PSF directly from a PDB-file without editing it.

12.10.2. Trajectory Handling. CHARMM can read trajectories that are contained in multiple files, and normally applies a number of checks to ensure that the set of files constitutes a valid contiguous trajectory with no overlaps or gaps. Sometimes it is desirable to override these checks, for instance, to analyze a set of independent replicate trajectories together to obtain overall statistics. These checks can be disabled, making it possible to mix files that are not contiguous or differ in a other ways (e.g., time step or coordinate saving frequency), as long as they use the same PSF.

Binary trajectories can be read automatically, irrespective of big-endian or little-endian format of the trajectory (and of the executable). The CHARMM file OPEN command also allows endianness to be specified with a keyword.

12.10.3. Time Series Analysis in the CORREL Module. Time series data can be mapped to specified interval, which is useful, e.g., to avoid spurious jumps in dihedral angles (MANTIM command). Time series of protein secondary structure content can also be extracted using CORREL.

12.10.4. Similarity Analysis of Snapshots from Trajectory Files. The RMSDYN command now allows different numbers of frames in two trajectories to be compared. Two new metrics have been added: the interatomic average coordinate difference without superposition (DIFF option), and the RMS distance (DRMS option), which compares interatomic distances in one structure with the corresponding distances in another structure, obviating the need for structural superposition. The latter is also available for single coordinate sets, and for time-series analysis in the CORREL module.

13. CONCLUDING DISCUSSION

Since its first publication in 1983,[4] CHARMM has been continuously developing as the need and demand for computational biophysics and biochemistry grew. The present review of the major developments since 2009[3] highlight improvements as well as new capability, many of which are uniquely available in CHARMM. This review may thereby serve as a guide for exploring new methods in addition to providing a broad overview of the current state of the art.

The advances also reflect changes in the research landscape at large. Faster simulation engines are needed as the molecular systems to study are becoming larger and also as the computer hardware continues to develop. With its modularity and flexibility, CHARMM now employs a number of engines either within the program or through APIs for external engines, which include DOMDEC, BLaDE, CHARMM/OpenMM, and the newly developed apoCHARMM (Section 2). Thus, the multicore/multithread scaling and speed of CHARMM should be comparable to those of other fast simulation engines presently available, while maintaining highest accuracy.

Accessibility is another practical issue, for which CHARMM is now readily available for academic and non-profit laboratories.[827] While the powerful CHARMM scripting language enables sophisticated tasks, a potential downside is the steep learning curve and difficulty in programming. This is being addressed through the development of pyCHARMM (Section 3.1). Its Python-based workflow also allows leveraging the capability of the Python language. Additionally, pyCHARMM is beginning to serve as a teaching platform from which the general principles and ideas of molecular modeling and biomolecular simulation can be taught. Preparation of the simulation system and the CHARMM script can also be done through CHARMM-GUI (Section 3.3).

Among other significant features of CHARMM are a wide range of docking and sampling methods (Section 4−Section 7) and various energy functions including implicit solvent and membranes, coarse graining, as well as a host of constraint capabilities (Section 8−Section 9). The QM/MM methods described in Section 11 are uniquely available in CHARMM. Likewise, CHARMM has distinct capabilities in system preparation, structure manipulation, and coordinate/trajectory analysis (Section 12). Finally, the ever-expanding CHARMM FF (Section 10) is becoming the *de facto* standard that is widely adopted in other simulation packages.

The extensive capabilities of CHARMM enable simulations and quantitative analyses of systems ranging from small molecules to large biomolecular assemblies and membrane systems at both atomistic and CG levels. Beyond studying small model systems, CHARMM is now increasingly used to tackle problems of practical importance that involve larger sizes, longer simulation times, and more extensive sampling, which will continue to grow with advances in computer hardware and methodologies. For the latter, CHARMM has been the testbed for new computational methods, thereby it stays on the forefront of biomolecular modeling and simulation with a fertile link to its developers and users. We anticipate CHARMM will continue to play an essential role for addressing current problems and also for opening new avenues of research in biomolecular systems.

AUTHOR INFORMATION

Corresponding Authors

Wonmuk Hwang − *Department of Biomedical Engineering, Texas A&M University, College Station, Texas 77843, United States; Department of Materials Science and Engineering and Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, United States; Center for AI and Natural Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea;* orcid.org/0000-0001-7514-3186; Email: hwm@tamu.edu

Charles L. Brooks, III − *Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-8149-5417; Email: brookscl@umich.edu

Bernard R. Brooks − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-3586-2730; Email: brb@nhlbi.nih.gov

Authors

Steven L. Austin − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States*

AQ

Arnaud Blondel − *Institut Pasteur, Université Paris Cité, CNRS UMR3825, Structural Bioinformatics Unit, F-75015 Paris, France;* orcid.org/0000-0001-9586-3025
Eric D. Boittier − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland*
Stefan Boresch − *Faculty of Chemistry, Department of Computational Biological Chemistry, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0002-2793-6656
Matthias Buck − *Department of Physiology and Biophysics, Case Western Reserve University, School of Medicine, Cleveland, Ohio 44106, United States*
Joshua Buckner − *Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States*
Amedeo Caflisch − *Department of Biochemistry, University of Zürich, CH-8057 Zürich, Switzerland;* orcid.org/0000-0002-2317-6792
Hao-Ting Chang − *Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan, ROC*
Xi Cheng − *Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China;* orcid.org/0000-0003-3735-645X
Yeol Kyo Choi − *Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, United States;* orcid.org/0000-0002-4218-7139
Jhih-Wei Chu − *Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, Institute of Molecular Medicine and Bioengineering, and Center for Intelligent Drug Systems and Smart Bio-devices (IDS^{2}B), National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan, ROC;* orcid.org/0000-0003-3842-2893
Michael F. Crowley − *Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, United States;* orcid.org/0000-0001-5163-9398
Qiang Cui − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; Department of Physics, Boston University, Boston, Massachusetts 02215, United States; Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, United States;* orcid.org/0000-0001-6214-5211
Ana Damjanovic − *Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, United States; Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, United States; Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0001-7635-2705
Yuqing Deng − *Shanghai R&D Center, DP Technology, Ltd., Shanghai 201210, China*
Mike Devereux − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* orcid.org/0000-0002-1561-1635
Xinqiang Ding − *Department of Chemistry, Tufts University, Medford, Massachusetts 02155, United States;* orcid.org/0000-0002-4598-8732
Michael F. Feig − *Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-9380-6422
Jiali Gao − *School of Chemical Biology & Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, Guangdong 518055, China; Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, Guangdong 518055, China; Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, United States;* orcid.org/0000-0003-0106-7154
David R. Glowacki − *CiTIUS Centro Singular de Investigación en Tecnoloxías Intelixentes da USC, 15705 Santiago de Compostela, Spain;* orcid.org/0000-0002-9608-3845
James E. Gonzales, II − *Department of Biomedical Engineering, Texas A&M University, College Station, Texas 77843, United States; Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0009-0009-7696-7561
Mehdi Bagerhi Hamaneh − *Department of Physiology and Biophysics, Case Western Reserve University, School of Medicine, Cleveland, Ohio 44106, United States;* Present Address: National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, United States; orcid.org/0009-0000-5222-889X
Edward D. Harder − *Schrodinger, Inc., New York, New York 10036, United States;* orcid.org/0000-0001-9213-787X
Ryan L. Hayes − *Department of Chemical and Biomolecular Engineering and Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, California 92697, United States;* orcid.org/0000-0003-1052-6391
Jing Huang − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China;* orcid.org/0000-0001-9639-2907
Yandong Huang − *College of Computer Engineering, Jimei University, Xiamen 361021, China;* orcid.org/0000-0002-1452-6383
Phillip S. Hudson − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States; Medicine Design, Pfizer Inc., Cambridge, Massachusetts 02139, United States*
Wonpil Im − *Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, United States;* orcid.org/0000-0001-5642-6041
Shahidul M. Islam − *Department of Chemistry, Delaware State University, Dover, Delaware 19901, United States;* orcid.org/0000-0001-5769-6844
Wei Jiang − *Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, United States;* orcid.org/0000-0001-7707-6247
Michael R. Jones − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-0542-3796
Silvan Käser − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* orcid.org/0000-0002-3641-8519
Fiona L. Kearns − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States;* orcid.org/0000-0002-5469-9035
Nathan R. Kern − *Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, United States;* orcid.org/0000-0003-0936-7274
Jeffery B. Klauda − *Department of Chemical and Biomolecular Engineering, Institute for Physical Science and*

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Technology, Biophysics Program, University of Maryland, College Park, Maryland 20742, United States; orcid.org/0000-0001-8725-1870

Themis Lazaridis − *Department of Chemistry, City College of New York, New York, New York 10031, United States;* orcid.org/0000-0003-4218-7590

Jinhyuk Lee − *Disease Target Structure Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Republic of Korea; Department of Bioinformatics, KRIBB School of Bioscience, University of Science and Technology, Daejeon 34141, Republic of Korea*

Justin A. Lemkul − *Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-6661-8653

Xiaorong Liu − *Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-7687-3604

Yun Luo − *Department of Biotechnology and Pharmaceutical Sciences, College of Pharmacy, Western University of Health Sciences, Pomona, California 91766, United States;* orcid.org/0000-0003-3581-754X

Alexander D. MacKerell, Jr. − *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, Maryland 21201, United States;* orcid.org/0000-0001-8287-6804

Dan T. Major − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel;* orcid.org/0000-0002-9231-0676

Markus Meuwly − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland; Department of Chemistry, Brown University, Providence, Rhode Island 02912, United States;* orcid.org/0000-0001-7930-8806

Kwangho Nam − *Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019, United States;* orcid.org/0000-0003-0723-7839

Lennart Nilsson − *Karolinska Institutet, Department of Biosciences and Nutrition, SE-14183 Huddinge, Sweden;* orcid.org/0000-0002-5067-6397

Victor Ovchinnikov − *Harvard University, Department of Chemistry and Chemical Biology, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0002-1793-2352

Emanuele Paci − *Dipartimento di Fisica e Astronomia, Universitá di Bologna, Bologna 40127, Italy;* orcid.org/0000-0002-4891-2768

Soohyung Park − *Department of Biological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015, United States;* orcid.org/0000-0002-4883-3031

Richard W. Pastor − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-2454-5131

Amanda R. Pittman − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States*

Carol Beth Post − *Borch Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907, United States;* orcid.org/0000-0002-5233-899X

Samarjeet Prasad − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Jingzhi Pu − *Department of Chemistry and Chemical Biology, Indiana University Indianapolis, Indianapolis, Indiana 46202, United States;* orcid.org/0000-0002-3042-335X

Yifei Qi − *School of Pharmacy, Fudan University, Shanghai 201203, China;* orcid.org/0000-0003-2853-7910

Thenmalarchelvi Rathinavelan − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana State 502284, India;* orcid.org/0000-0002-1142-0583

Daniel R. Roe − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-5834-2447

Benoit Roux − *Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-5254-2712

Christopher N. Rowley − *Department of Chemistry, Carleton University, Ottawa, Ontario K1S 5B6, Canada;* orcid.org/0000-0002-0205-952X

Jana Shen − *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, Maryland 21201, United States;* orcid.org/0000-0002-3234-0769

Andrew C. Simmonett − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-5921-9272

Alexander J. Sodt − *Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, United States;* orcid.org/0000-0002-5570-8212

Kai Töpfer − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* orcid.org/0000-0002-4650-9641

Meenu Upadhyay − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* orcid.org/0000-0003-3802-0835

Arjan van der Vaart − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States;* orcid.org/0000-0002-8950-1850

Luis Itza Vazquez-Salazar − *Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* orcid.org/0000-0001-6347-5108

Richard M. Venable − *Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Luke C. Warrensford − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States;* orcid.org/0000-0002-9206-5121

H. Lee Woodcock − *Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States;* orcid.org/0000-0003-3539-273X

Yujin Wu − *Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States*

Martin Karplus − *Harvard University, Department of Chemistry and Chemical Biology, Cambridge, Massachusetts 02138, United States; Laboratoire de Chimie Biophysique, ISIS, Université de Strasbourg, 67000 Strasbourg, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.4c04100

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Author Contributions

M. Karplus provided the overall guidance of the CHARMM Development Project. W. Hwang assembled the manuscript; W. Hwang, C.L. Brooks, III, and B.R. Brooks edited the manuscript; J.E. Gonzales, II and W. Hwang edited the references. Sections 1, and 13: W. Hwang, C.L. Brooks, III, and B.R. Brooks. Sections 2.1, 2.2, 2.3, 3.1, 3.2, 4.1, and 5.1: R.L. Hayes, J. Buckner, X. Liu, Y. Wu, X. Ding, M.F. Crowley, and C.L. Brooks, III. Sections 2.4 and 12.4: S. Prasad and B.R. Brooks. Section 3.3: Y. Qi, X. Cheng Y. Choi, N.R. Kern, and W. Im. Sections 4.2, 8.4, 11.1, 11.2, 11.3, 11.4, and 11.8: K. Nam, J. Pu, D.T. Major, and J. Gao. Section 4.3: L.C. Warrensford, A.R. Pittman, F.L. Kearns, and H.L. Woodcock. Section 5.2: B. Roux, W. Jiang, and Y. Deng. Section 5.3: A. Blondel. Section 6.1: J. Shen and Y. Huang. Section 6.2: A. Damjanovic and B.R. Brooks. Sections 6.3, 8.7.1, and 8.7.2: T. Lazaridis. Section 7.1: D.R. Roe and B.R. Brooks. Section 7.2: A. van der Vaart. Sections 7.3 (introduction), 7.3.1, 7.3.2, 7.3.3, and 7.3.4: V. Ovchinnikov and M. Karplus. Sections 7.3.5 and 7.3.6: B. Roux, Y. Luo, and E.D. Harder. Sections 7.4 and 12.9: C. Post. Section 7.5: H.T. Chang and J.W. Chu. Sections 7.6 and 11.9: E. Paci and D.R. Glowacki. Section 7.7: S.L. Austin and H.L. Woodcock. Section 8.1: M. Devereux, E.D. Boittier, K. Töpfer, and M. Meuwly. Section 8.2: K. Töpfer, S. Käser, L.I. Vazquez-Salazar, and M. Meuwly. Section 8.3: J. Huang, A. MacKerell, Jr., and B.R. Brooks. Section 8.5: A.C. Simmonett, B.R. Brooks, and R.W. Pastor. Section 8.6: A. Caflisch. Sections 8.7.3 and 8.8: M. Feig. Sections 9.1 and 9.2: W. Im and J. Lee. Section 9.3: W. Im and T. Rathinavelan. Sections 9.4 and 12.10: L. Nilsson. Section 10: A.D. MacKerell, Jr., J. Huang, J.B. Klauda, J.A. Lemkul, L. Nilsson, R.W. Pastor, and B. Roux. Section 11.5: Q. Cui. Section 11.6: M.R. Jones and B.R. Brooks. Section 11.7: B. Roux and C.N. Rowley. Section 11.10: M. Upadhyay, and M. Meuwly. Section 11.11: S. Boresch, H.L. Woodcock, P.S. Hudson, and F.L. Kearns. Section 12.1: B. Roux and S.M. Islam. Section 12.2: D.R. Roe, R.M. Venable, R.W. Pastor, and B.R. Brooks. Section 12.3: S. Park, A.J. Sodt, W. Im, and R.W. Pastor. Section 12.5: M. Buck and M.B. Hamaneh. Sections 12.6, 12.7, and 12.8: W. Hwang and J.E. Gonzales, II.

Notes

The authors declare the following competing financial interest(s): Alexander D. MacKerell, Jr. is Co-founder and CSO of SilcsBio LLC.

■ ACKNOWLEDGMENTS

This work was supported by various funding agencies including the U.S. National Institutes of Health (NIH) and the U.S. National Science Foundation (NSF). The following separate acknowledgments are in response to requests from the authors: BR Brooks and RW Pastor: Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute (BR Brooks: ZIA HL001051 and ZIA HL001052; RW Pastor: ZIA HL000340). CL Brooks: NIH R35GM137058. A Caflisch: Swiss National Science Foundation (grant number 310030-212195). JW Chu: National Science and Technology Council of Taiwan (112-2113-M-A49-018-MY3). Q Cui: NIH R35GM141930. M Feig: NIH R35GM126948. J Gao: NIH R01GM046736. D Glowacki: European Research Council Consolidator Grant NANOVR 866559. MB Hamaneh: Intramural program of the National Library of Medicine at the NIH. W Im: NSF MCB-211172 and NIH R01GM138472. W Jiang: The Office of Science, U.S. Department of Energy (DE-AC02-06CH11357). T Lazaridis: NSF MCB-1855942. JA Lemkul: R35GM133754. AD MacKerell, Jr.: NIH R35GM131710. M Meuwly: The Swiss National Science Foundation (grants 215088 and 219779), the NCCR MUST, and the University of Basel. K Nam and DT Major: NIH R01GM132481 (K Nam) and NIH R21GM148895 (K Nam and DT Major). CB Post: NIH R01GM039478. J Pu: NIH R01GM135392. J Shen: NIH R35GM148261. A van der Vaart: NSF MCB-1919096. HL Woodcock: NIH R01GM129519, R01GM143810, and R01GM140316.

■ REFERENCES

(1) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585−90.
(2) Macuglia, D.; Roux, B.; Ciccotti, G. The emergence of protein dynamics simulations: how computational statistical mechanics met biochemistry. *Euro. Phys. J. H* **2022**, *47*, 13.
(3) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(4) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(5) Harris, J. A.; Liu, R.; de Oliveira, V. M.; Vázquez-Montelongo, E. A.; Henderson, J. A.; Shen, J. GPU-accelerated all-atom particle-mesh Ewald continuous constant pH molecular dynamics in Amber. *J. Chem. Theory Comp.* **2022**, *18*, 7510−7527.
(6) Aho, N.; Buslaev, P.; Jansen, A.; Bauer, P.; Groenhof, G.; Hess, B. Scalable constant pH molecular dynamics in GROMACS. *J. Chem. Theory Comp.* **2022**, *18*, 6148−6160.
(7) Jiang, W.; Chipot, C.; Roux, B. Computing relative binding affinity of ligands to receptor: an effective hybrid single-dual-topology free-energy perturbation approach in NAMD. *J. Chem. Inf. and Mod.* **2019**, *59*, 3794−3802.
(8) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(9) Buckner, J.; Liu, X.; Chakravorty, A.; Wu, Y.; Cervantes, L. F.; Lai, T. T.; Brooks, C. L., III. pyCHARMM: embedding CHARMM functionality in a python framework. *J. Chem. Theory Comp.* **2023**, *19*, 3752−3762.
(10) Chakravorty, A.; Hussain, A.; Cervantes, L. F.; Lai, T.; Brooks, C. L., III. Exploring the limits of small molecule force fields through predictions of solvation free energy of small molecules. *J. Chem. Inf. and Mod.* **2024**, *64*, 4089−4101.
(11) Ding, X.; Vilseck, J. Z.; Brooks, C. L., III. Fast solver for large scale multistate Bennett acceptance ratio equations. *J. Chem. Theory Comp.* **2019**, *15*, 799−802.
(12) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(13) Ding, X.; Wu, Y.; Wang, Y.; Vilseck, J. Z.; Brooks, C. L., III. Accelerated CDOCKER with GPUs, parallel simulated annealing, and fast Fourier transforms. *J. Chem. Theory Comp.* **2020**, *16*, 3910−3919.
(14) Wu, Y.; Brooks, C. L., III. Flexible CDOCKER: hybrid searching algorithm and scoring function with side chain conformational entropy. *J. Chem. Inf. and Mod.* **2021**, *61*, 5535−5549.
(15) Wu, Y.; Brooks, C. L., III. Covalent docking in CDOCKER. *J. Comput. Aided. Mol. Des.* **2022**, *36*, 563−574.
(16) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−616.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(17) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. and Mod.* **2012**, *52*, 3144−3154.
(18) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. and Mod.* **2012**, *52*, 3155−3168.
(19) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comp.* **2020**, *16*, 528−552.
(20) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153*, 114502.
(21) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(22) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331−W336.
(23) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., Jr. An empirical polarizable force field based on the classical Drude oscillator model: development history and recent applications. *Chem. Rev.* **2016**, *116*, 4983−5013.
(24) Hayes, R. L.; Buckner, J.; Brooks, C. L., III. BLaDE: a basic lambda dynamics engine for GPU accelerated molecular dynamics free energy calculations. *J. Chem. Theory Comp.* **2021**, *17*, 6799−6807.
(25) Hynninen, A. P.; Crowley, M. F. New faster CHARMM molecular dynamics engine. *J. Comput. Chem.* **2014**, *35*, 406−413.
(26) Knight, J. L.; Brooks, C. L., III. Multisite λ dynamics for simulated structure-activity relationship studies. *J. Chem. Theory Comp.* **2011**, *7*, 2728−2739.
(27) Vilseck, J. Z.; Armacost, K. A.; Hayes, R. L.; Goh, G. B.; Brooks, C. L., III. Predicting binding free energies in a large combinatorial chemical space using multisite λ dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 3328−3332.
(28) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L., III. Automated, accurate, and scalable relative protein-ligand binding free-energy calculations using lambda dynamics. *J. Chem. Theory Comp.* **2020**, *16*, 7895−7914.
(29) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J. Chem. Theory Comp.* **2012**, *8*, 36−46.
(30) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. Towards accurate prediction of protonation equilibrium of nucleic acids. *J. Phys. Chem. Lett.* **2013**, *4*, 760−766.
(31) Armacost, K. A.; Goh, G. B.; Brooks, C. L., III. Biasing potential replica exchange multisite λ-dynamics for efficient free energy calculations. *J. Chem. Theory Comp.* **2015**, *11*, 1267−1277.
(32) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(33) Chow, K.-H.; Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput. Phys. Commun.* **1995**, *91*, 283−289.
(34) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288−294.
(35) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald - an NLog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(36) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667−683.
(37) Christ, C. D.; van Gunsteren, W. F. Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J. Chem. Phys.* **2007**, *126*, 184110.
(38) Christ, C. D.; van Gunsteren, W. F. Multiple free energies from a single simulation: extending enveloping distribution sampling to nonoverlapping phase-space distributions. *J. Chem. Phys.* **2008**, *128*, 174112.
(39) Christ, C. D.; Van, W. F. Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. *J. Comput. Chem.* **2009**, *30*, 1664−1679.
(40) Lee, J.; Miller, B. T.; Damjanović, A.; Brooks, B. R. Constant pH molecular dynamics in explicit solvent with enveloping distribution sampling and Hamiltonian exchange. *J. Chem. Theory Comp.* **2014**, *10*, 2738−2750.
(41) Lee, J.; Miller, B. T.; Damjanović, A.; Brooks, B. R. Enhancing constant-pH simulation in explicit solvent with a two-dimensional replica exchange method. *J. Chem. Theory Comp.* **2015**, *11*, 2560−2574.
(42) Dolan, E. A.; Venable, R. M.; Pastor, R. W.; Brooks, B. R. Simulations of membranes and other interfacial systems using P21 and Pc periodic boundary conditions. *Biophys. J.* **2002**, *82*, 2317−2325.
(43) Prasad, S.; Simmonett, A. C.; Meana-Paneda, R.; Brooks, B. R. The extended eighth-shell method for periodic boundary conditions with rotational symmetry. *J. Comput. Chem.* **2021**, *42*, 1373−1383.
(44) Park, S.; Rice, A.; Im, W.; Pastor, R. W. Spontaneous curvature generation by peptides in asymmetric bilayers. *J. Comput. Chem.* **2023**, *45*, 512−522.
(45) Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; et al. Best practices for alchemical free energy calculations. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378.
(46) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(47) Wu, X.; Brooks, B. R. A double exponential potential for van der Waals interaction. *AIP Adv.* **2019**, *9*, 065304.
(48) Hynninen, A.-P. ap-hynninen/GSE. 2017.
(49) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(50) Martins de Oliveira, V.; Liu, R.; Shen, J. Constant pH molecular dynamics simulations: current status and recent applications. *Curr. Op. Struct. Biol.* **2022**, *77*, 102498.
(51) Brunger, A. T.; Campbell, R. L.; Clore, G. M.; Gronenborn, A. M.; Karplus, M.; Petsko, G. A.; Teeter, M. M. Solution of a protein crystal structure with a model obtained from NMR interproton distance restraints. *Science* **1987**, *235*, 1049−1053.
(52) Raman, E. P.; Yu, W.; Guvench, O.; Mackerell, A. D., Jr. Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J. Chem. Inf. and Mod.* **2011**, *51*, 877−896.
(53) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. and Mod.* **2020**, *60*, 3408−3415.
(54) Unke, O. T.; Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comp.* **2019**, *15*, 3678−3693.
(55) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER-a CHARMm-based Md docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(56) Gagnon, J. K.; Law, S. M.; Brooks, C. L., III. Flexible CDOCKER: development and application of a pseudo-explicit structure-based docking method within CHARMM. *J. Comput. Chem.* **2016**, *37*, 753−762.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(57) Buckner, J.; Liu, X.; Chakravorty, A.; Wu, Y.; Cervantes, L. F.; Lai, T. T.; Brooks, C. L., III. pyCHARMM-Workshop (accessed August 5, 2024); https://github.com/BrooksResearchGroup-UM/pyCHARMM-Workshop.
(58) Hayes, L., Ryan, Vilseck, J. Z., Buckner, J., Liu, X., Chakravorty, A., Cervantes, L. F., Lai, T. T., Brooks, C. L., III. MSLD-Workshop 2024, https://github.com/BrooksResearchGroup-UM/MSLD-Workshop.
(59) Jo, S.; Kim, T.; Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2007**, *2*, No. e880.
(60) Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S. J.; Patel, D. S.; Beaven, A. H.; Lee, K. I.; Rui, H.; Park, S.; et al. CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.* **2017**, *38*, 1114−1124.
(61) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **2009**, *97*, 50−58.
(62) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Davila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35*, 1997−2004.
(63) Kunzmann, P.; Muller, T. D.; Greil, M.; Krumbach, J. H.; Anter, J. M.; Bauer, D.; Islam, F.; Hamacher, K. Biotite: new tools for a versatile Python bioinformatics library. *BMC Bioinf* **2023**, *24*, 236.
(64) Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinf* **2009**, *25*, 1422−3.
(65) Xu, L., Ziqiao; Brooks, C. L., III. crimm (accessed August 5, 2024); https://github.com/BrooksResearchGroup-UM/crimm.
(66) Rose, A. S.; Bradley, A. R.; Valasatava, Y.; Duarte, J. M.; Prlic, A.; Rose, P. W. NGL viewer: web-based molecular graphics for large complexes. *Bioinf* **2018**, *34*, 3755−3758.
(67) Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.; Craig, P. A.; Crichlow, G. V.; Dalenberg, K.; Duarte, J. M.; et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **2023**, *51*, D488−D508.
(68) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590−596.
(69) Westbrook, J. D.; Young, J. Y.; Shao, C.; Feng, Z.; Guranovic, V.; Lawson, C. L.; Vallat, B.; Adams, P. D.; Berrisford, J. M.; Bricogne, G.; et al. PDBx/mmCIF ecosystem: foundational semantic tools for structural biology. *J. Mol. Biol.* **2022**, *434*, 167599.
(70) Sondergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comp.* **2011**, *7*, 2284−95.
(71) Landrum, G.; Tosco, P.; Kelley, B.; Rodriguez, R.; Cosgrove, D.; Vianello, R.; Schneider, N.; Kawashima, E.; Jones, G.; Nealschneider, D.; et al. RDKit (accessed August 5, 2024);.
(72) Boothroyd, S.; Behara, P. K.; Madin, O. C.; Hahn, D. F.; Jang, H.; Gapsys, V.; Wagner, J. R.; Horton, J. T.; Dotson, D. L.; Thompson, M. W.; et al. Development and benchmarking of open force field 2.0.0: the sage small molecule force field. *J. Chem. Theory Comp.* **2023**, *19*, 3251−3275.
(73) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. and Mod.* **2021**, *61*, 3891−3898.
(74) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(75) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. 2006; p 84.
(76) Jung, J.; Mori, T.; Kobayashi, C.; Matsunaga, Y.; Yoda, T.; Feig, M.; Sugita, Y. GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 310−323.
(77) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(78) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
(79) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Henin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
(80) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardere, L.; Schnieders, M. J.; Piquemal, J. P.; Ren, P.; Ponder, J. W. Tinker 8: software tools for molecular design. *J. Chem. Theory Comp.* **2018**, *14*, 5273−5289.
(81) Woo, H.; Park, S. J.; Choi, Y. K.; Park, T.; Tanveer, M.; Cao, Y.; Kern, N. R.; Lee, J.; Yeom, M. S.; Croll, T. I.; et al. Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *J. Phys. Chem. B* **2020**, *124*, 7128−7137.
(82) Pogozheva, I. D.; Armstrong, G. A.; Kong, L.; Hartnagel, T. J.; Carpino, C. A.; Gee, S. E.; Picarello, D. M.; Rubin, A. S.; Lee, J.; Park, S.; et al. Comparative molecular dynamics simulation studies of realistic eukaryotic, prokaryotic, and archaeal membranes. *J. Chem. Inf. and Mod.* **2022**, *62*, 1036−1051.
(83) Pogozheva, I. D.; Cherepanov, S.; Park, S. J.; Raghavan, M.; Im, W.; Lomize, A. L. Structural modeling of cytokine-receptor-JAK2 signaling complexes using AlphaFold multimer. *J. Chem. Inf. and Mod.* **2023**, *63*, 5874−5895.
(84) Cao, Y.; Choi, Y. K.; Frank, M.; Woo, H.; Park, S. J.; Yeom, M. S.; Seok, C.; Im, W. Dynamic interactions of fully glycosylated SARS-CoV-2 spike protein with various antibodies. *J. Chem. Theory Comp.* **2021**, *17*, 6559−6569.
(85) Choi, Y. K.; Cao, Y.; Frank, M.; Woo, H.; Park, S. J.; Yeom, M. S.; Croll, T. I.; Seok, C.; Im, W. Structure, dynamics, receptor binding, and antibody binding of the fully glycosylated full-length SARS-CoV-2 spike protein in a viral membrane. *J. Chem. Theory Comp.* **2021**, *17*, 2479−2487.
(86) Qi, Y.; Cheng, X.; Han, W.; Jo, S.; Schulten, K.; Im, W. CHARMM-GUI PACE CG Builder for solution, micelle, and bilayer coarse-grained simulations. *J. Chem. Inf. and Mod.* **2014**, *54*, 1003−1009.
(87) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comp.* **2016**, *12*, 405−413.
(88) Lee, J.; Hitzenberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W. CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* **2020**, *153*, 035103.
(89) Cheng, X.; Jo, S.; Lee, H. S.; Klauda, J. B.; Im, W. CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems. *J. Chem. Inf. and Mod.* **2013**, *53*, 2171−2180.
(90) Qi, Y.; Cheng, X.; Lee, J.; Vermaas, J. V.; Pogorelov, T. V.; Tajkhorshid, E.; Park, S.; Klauda, J. B.; Im, W. CHARMM-GUI HMMM builder for membrane simulations with the highly mobile membrane-mimetic model. *Biophys. J.* **2015**, *109*, 2012−2022.
(91) Lee, J.; Patel, D. S.; Stahle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; et al. CHARMM-GUI

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

membrane builder for complex biological membrane simulations with glycolipids and lipoglycans. *J. Chem. Theory Comp.* **2019**, *15*, 775−786.
(92) Qi, Y.; Lee, J.; Klauda, J. B.; Im, W. CHARMM-GUI nanodisc builder for modeling and simulation of various nanodisc systems. *J. Comput. Chem.* **2019**, *40*, 893−899.
(93) Park, S.; Choi, Y. K.; Kim, S.; Lee, J.; Im, W. CHARMM-GUI membrane builder for lipid nanoparticles with ionizable cationic lipids and PEGylated lipids. *J. Chem. Inf. and Mod.* **2021**, *61*, 5192−5202.
(94) Feng, S.; Park, S.; Choi, Y. K.; Im, W. CHARMM-GUI membrane builder: past, current, and future developments and applications. *J. Chem. Theory Comp.* **2023**, *19*, 2161−2185.
(95) Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; et al. CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues. *Adv. Prot. Chem. Struct. Biol.* **2014**, *96*, 235−265.
(96) Park, S. J.; Kern, N.; Brown, T.; Lee, J.; Im, W. CHARMM-GUI PDB manipulator: various PDB structural modifications for biomolecular modeling and simulation. *J. Mol. Biol.* **2023**, *435*, 167995.
(97) Jo, S.; Jiang, W.; Lee, H. S.; Roux, B.; Im, W. CHARMM-GUI ligand binder for absolute binding free energy calculations and its application. *J. Chem. Inf. and Mod.* **2013**, *53*, 267−277.
(98) Kim, S.; Lee, J.; Jo, S.; Brooks, C. L., III; Lee, H. S.; Im, W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **2017**, *38*, 1879−1886.
(99) Kim, S.; Oshima, H.; Zhang, H.; Kern, N. R.; Re, S.; Lee, J.; Roux, B.; Sugita, Y.; Jiang, W.; Im, W. CHARMM-GUI free energy calculator for absolute and relative ligand solvation and binding free energy simulations. *J. Chem. Theory Comp.* **2020**, *16*, 7207−7218.
(100) Guterres, H.; Park, S. J.; Zhang, H.; Im, W. CHARMM-GUI LBS finder & refiner for ligand binding site prediction and refinement. *J. Chem. Inf. and Mod.* **2021**, *61*, 3744−3751.
(101) Guterres, H.; Park, S. J.; Cao, Y.; Im, W. CHARMM-GUI ligand designer for template-based virtual ligand design in a binding site. *J. Chem. Inf. and Mod.* **2021**, *61*, 5336−5342.
(102) Zhang, H.; Kim, S.; Giese, T. J.; Lee, T. S.; Lee, J.; York, D. M.; Im, W. CHARMM-GUI free energy calculator for practical ligand binding free energy simulations with AMBER. *J. Chem. Inf. and Mod.* **2021**, *61*, 4145−4151.
(103) Guterres, H.; Park, S.-J.; Zhang, H.; Perone, T.; Kim, J.; Im, W. CHARMM-GUI high-throughput simulator for efficient evaluation of protein-ligand interactions with different force fields. *Protein Sci.* **2022**, *31*, No. e4413.
(104) Guterres, H.; Im, W. CHARMM-GUI-based induced fit docking workflow to generate reliable protein-ligand binding modes. *J. Chem. Inf. and Mod.* **2023**, *63*, 4772−4779.
(105) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D., Jr.; Im, W. Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **2011**, *32*, 3135−3141.
(106) Park, S. J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. Glycan reader is improved to recognize most sugar types and chemical modifications in the protein data bank. *Bioinf* **2017**, *33*, 3051−3057.
(107) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI glycan modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiol* **2019**, *29*, 320−331.
(108) Jo, S.; Vargyas, M.; Vasko-Szedlar, J.; Roux, B.; Im, W. PBEQ-solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* **2008**, *36*, W270−W275.
(109) Qi, Y.; Ingolfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI martini maker for coarse-grained simulations with the martini force field. *J. Chem. Theory Comp.* **2015**, *11*, 4486−4494.
(110) Hsu, P. C.; Bruininks, B. M. H.; Jefferies, D.; Cesar Telles de Souza, P.; Lee, J.; Patel, D. S.; Marrink, S. J.; Qi, Y.; Khalid, S.; Im, W. CHARMM-GUI martini maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides. *J. Comput. Chem.* **2017**, *38*, 2354−2363.
(111) Qi, Y.; Lee, J.; Cheng, X.; Shen, R.; Islam, S. M.; Roux, B.; Im, W. CHARMM-GUI DEER facilitator for spin-pair distance distribution calculations and preparation of restrained-ensemble molecular dynamics simulations. *J. Comput. Chem.* **2020**, *41*, 415−420.
(112) Choi, Y. K.; Park, S. J.; Park, S.; Kim, S.; Kern, N. R.; Lee, J.; Im, W. CHARMM-GUI polymer builder for modeling and simulation of synthetic polymers. *J. Chem. Theory Comp.* **2021**, *17*, 2431−2443.
(113) Gao, Y.; Lee, J.; Smith, I. P. S.; Lee, H.; Kim, S.; Qi, Y.; Klauda, J. B.; Widmalm, G.; Khalid, S.; Im, W. CHARMM-GUI supports hydrogen mass repartitioning and different protonation states of phosphates in lipopolysaccharides. *J. Chem. Inf. and Mod.* **2021**, *61*, 831−839.
(114) Choi, Y. K.; Kern, N. R.; Kim, S.; Kanhaiya, K.; Afshar, Y.; Jeon, S. H.; Jo, S.; Brooks, B. R.; Lee, J.; Tadmor, E. B.; et al. CHARMM-GUI nanomaterial modeler for modeling and simulation of nanomaterial systems. *J. Chem. Theory Comp.* **2022**, *18*, 479−493.
(115) Kognole, A. A.; Lee, J.; Park, S. J.; Jo, S.; Chatterjee, P.; Lemkul, J. A.; Huang, J.; MacKerell, A. D., Jr.; Im, W. CHARMM-GUI Drude prepper for molecular dynamics simulation using the classical Drude polarizable force field. *J. Comput. Chem.* **2022**, *43*, 359−375.
(116) Zhang, H.; Kim, S.; Im, W. Practical guidance for consensus scoring and force field selection in protein-ligand binding free energy simulations. *J. Chem. Inf. and Mod.* **2022**, *62*, 6084−6093.
(117) Lee, K. I.; Jo, S.; Rui, H.; Egwolf, B.; Roux, B.; Pastor, R. W.; Im, W. Web interface for Brownian dynamics simulation of ion transport and its applications to beta-barrel pores. *J. Comput. Chem.* **2012**, *33*, 331−339.
(118) Qi, Y.; Lee, J.; Singharoy, A.; McGreevy, R.; Schulten, K.; Im, W. CHARMM-GUI MDFF/xMDFF Utilizer for molecular dynamics flexible fitting simulations in various environments. *J. Phys. Chem. B* **2017**, *121*, 3718−3723.
(119) Suh, D.; Feng, S.; Lee, H.; Zhang, H.; Park, S. J.; Kim, S.; Lee, J.; Choi, S.; Im, W. CHARMM-GUI enhanced sampler for various collective variables and enhanced sampling methods. *Protein Sci.* **2022**, *31*, No. e4446.
(120) Wang, K. W.; Lee, J.; Zhang, H.; Suh, D.; Im, W. CHARMM-GUI implicit solvent modeler for various generalized Born models in different simulation programs. *J. Phys. Chem. B* **2022**, *126*, 7354−7364.
(121) Kondo, D.; Taufer, M.; Brooks, C. L., III; Casanova, H.; Chien, A. A. Characterizing and evaluating desktop grids: An empirical study. *18th International Parallel and Distributed Processing Symposium, 2004. Proceedings*. 2004, p 26.
(122) Taufer, M.; Kerstens, A.; Estrada, T.; Flores, D. A.; Zamudio, R.; Teller, P. J.; Armen, R. S.; Brooks, C. L., III. Moving volunteer computing towards knowledge-constructed, dynamically-adaptive modeling and scheduling. *International Parallel and Distributed Processing Symposium* 2007, 1−8.
(123) Taufer, M.; An, C.; Kerstens, A.; Brooks, C. L., III. Predictor@ Home: a "protein structure prediction supercomputer" based on global computing. *IEEE Transac. Paral. Distrib. Sys.* **2006**, *17*, 786−796.
(124) Taufer, M.; Armen, R.; Chen, J.; Teller, P.; Brooks, C. L., III. Computational multiscale modeling in protein−ligand docking. *IEEE Eng. Med. Biol. Mag.* **2009**, *28*, 58−69.
(125) Rahaman, O.; Estrada, T. P.; Doren, D. J.; Taufer, M.; Brooks, C. L., III; Armen, R. S. Evaluation of several two-step scoring functions based on linear interaction energy, effective ligand size, and empirical pair potentials for prediction of protein-ligand binding geometry and free energy. *J. Chem. Inf. and Mod.* **2011**, *51*, 2047−2065.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(126) Armen, R. S.; Chen, J.; Brooks, C. L., III. An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comp.* **2009**, *5*, 2909−2923.
(127) Ahmed, F.; Brooks, C. L., III. FASTDock: a pipeline for allosteric drug discovery. *J. Chem. Inf. and Mod.* **2023**, *63*, 7219−7227.
(128) Horton, J. T.; Boothroyd, S.; Wagner, J.; Mitchell, J. A.; Gokey, T.; Dotson, D. L.; Behara, P. K.; Ramaswamy, V. K.; Mackey, M.; Chodera, J. D.; et al. Open Force Field BespokeFit: automating bespoke torsion parametrization at scale. *J. Chem. Inf. and Mod.* **2022**, *62*, 5622−5633.
(129) Dominy, B. N.; Brooks, C. L., III. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(130) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(131) Im, W.; Lee, M. S.; Brooks, C. L., III. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(132) Haberthür, U.; Caflisch, A. FACTS: fast analytical continuum treatment of solvation. *J. Comput. Chem.* **2008**, *29*, 701−715.
(133) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **2017**, *549*, 502−506.
(134) Chang, Z.; Ansbacher, T.; Zhang, L.; Yang, Y.; Ko, T.-P.; Zhang, G.; Liu, W.; Huang, J.-W.; Dai, L.; Guo, R.-T.; et al. Crystal structure of Lepi, a multifunctional SAM-dependent enzyme which catalyzes pericyclic reactions in leporin biosynthesis. *Org. Biomol. Chem.* **2019**, *17*, 2070−2076.
(135) Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. EnzyDock: protein-ligand docking of multiple reactive states along a reaction coordinate in enzymes. *J. Chem. Theory Comp.* **2019**, *15*, 5116−5134.
(136) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III. Assessing search strategies for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1623−1631.
(137) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III. Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1612−1622.
(138) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Chem. Phys.* **1977**, *23*, 187−199.
(139) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free energy of conformational transition paths in biomolecules: the string method and its application to myosin VI. *J. Chem. Phys.* **2011**, *134*, 085103.
(140) Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200−212.
(141) Demapan, D.; Kussman, J.; Ochsenfeld, C.; Cui, Q. Factors that determine the variation of equilibrium and kinetic properties of QM/MM enzyme simulations: QM region, conformation and boundary condition. *J. Chem. Theory Comp.* **2022**, *18*, 2530−2542.
(142) Raz, K.; Driller, R.; Dimos, N.; Ringel, M.; Brück, T.; Loll, B.; Major, D. T. The impression of a nonexisting catalytic effect: the role of CotB2 in guiding the complex biosynthesis of Cyclooctat-9-en-7-ol. *J. Am. Chem. Soc.* **2020**, *142*, 21562−21574.
(143) Hu, J.; Ma, A.; Dinner, A. R. Monte Carlo simulations of biomolecules: the mc module in CHARMM. *J. Comput. Chem.* **2006**, *27*, 203−216.
(144) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Meth.* **2017**, *14*, 71−73.
(145) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(146) Hudson, P. S.; Aviat, F.; Meana-Pañeda, R.; Warrensford, L.; Pollard, B. C.; Prasad, S.; Jones, M. R.; Woodcock III, H. L.; Brooks, B. R. Obtaining QM/MM binding free energies in the SAMPL8 drugs of abuse challenge: indirect approaches. *J. Comput. Aided Mol. Des.* **2022**, *36*, 263−277.
(147) Ojeda-May, P.; Nam, K. Acceleration of semiempirical QM/MM methods through message passing interface (MPI), hybrid MPI/Open Mmltiprocessing, and self-consistent field accelerator implementations. *J. Chem. Theory Comp.* **2017**, *13*, 3525−3536.
(148) Woodcock III, H. L.; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; Brooks, B. R. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(149) Gupta, P. K.; Tarannam, N.; Zev, S.; Major, D. T. Multistate multiscale docking study of the hydrolysis of toxic nerve agents by phosphotriesterase. *Electron. Struct.* **2023**, *5*, 035003.
(150) Raz, K.; Levi, S.; Gupta, P. K.; Major, D. T. Enzymatic control of product distribution in terpene synthases: insights from multiscale simulations. *Curr. Opin. Biotechnol.* **2020**, *65*, 248−258.
(151) Zev, S.; Raz, K.; Schwartz, R.; Tarabeh, R.; Gupta, P. K.; Major, D. T. Benchmarking the ability of common docking programs to correctly reproduce and score binding modes in SARS-CoV-2 protease Mpro. *J. Chem. Inf. and Mod.* **2021**, *61*, 2957−2966.
(152) Koshland, D. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 98−104.
(153) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209−25.
(154) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534−553.
(155) Vankayala, S.; Warrensford, L. C.; Pittman, A.; Pollard, B.; Kearns, F.; Larkin, J.; Woodcock, H. L. CIFDock: A novel CHARMM-based flexible receptor−flexible ligand docking protocol. *J. Comput. Chem.* **2022**, *43*, 84−95.
(156) O'Boyle, N. M.; Banck, M.; James, C. A.; Morely, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.
(157) Wu, X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512−518.
(158) Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Struct. Dyn.* **1991**, *8*, 1267−89.
(159) Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270−W277.
(160) Lee, M. S.; Feig, M.; Salsbury Jr, F. R.; Brooks, C. L., III. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24*, 1348−1356.
(161) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. method. *J. Comput. Chem.* **1989**, *10*, 209−220.
(162) Elstner, M. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316−325.
(163) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **2001**, *105*, 569−585.
(164) Gaus, M.; Cui, Q.; Elstner, M. DFTB-3rd: extension of the self-consistent-charge density-functional tight-binding method SCC-DFTB. *J. Chem. Theory Comp.* **2011**, *7*, 931−948.
(165) Nam, K. Acceleration of semiempirical quantum mechanical calculations by extended Lagrangian molecular dynamics approach. *J. Chem. Theory Comp.* **2013**, *9*, 3393−3403.
(166) Toledo Warshaviak, D.; Golan, G.; Borrelli, K. W.; Zhu, K.; Kalid, O. Structure-based virtual screening approach for discovery of covalently bound ligands. *J. Chem. Inf. and Mod.* **2014**, *54*, 1941−1950.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(167) Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E. Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. *J. Chem. Inf. and Mod.* **2014**, *54*, 1932−1940.
(168) Wang, L.; Chambers, J.; Abel, R. *Biomolecular Simulations*; Meth. *Mol. Biol.* **2019**, *2022*, 201−232.
(169) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **2020**, *11*, 1140−1152.
(170) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angew. Chem.* **2016**, *55*, 7364−7368.
(171) Hayes, R. L.; Nixon, C. F.; Marqusee, S.; Brooks, C. L., III. Selection pressures on evolution of ribonuclease H explored with rigorous free-energy-based design. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, No. e2312029121.
(172) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., III. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *PROTEINS: Struct., Func. and Bioinf.* **2014**, *82*, 1319−1331.
(173) Kong, X.; Brooks, C. L., III. Lambda-Dynamics: a new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(174) Thiel, A. C.; Speranza, M. J.; Jadhav, S.; Stevens, L. L.; Unruh, D. K.; Ren, P.; Ponder, J. W.; Shen, J.; Schnieders, M. J. Constant-pH Simulations with the Polarizable Atomic Multipole AMOEBA Force Field. *J. Chem. Theory Comp.* **2024**, *20*, 2921−2933.
(175) Hayes, R. L.; Armacost, K. A.; Vilseck, J. Z.; Brooks, C. L., III. Adaptive landscape flattening accelerates sampling of alchemical space in multisite λ dynamics. *J. Phys. Chem. B* **2017**, *121*, 3626−3635.
(176) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(177) Hayes, R. L.; Vilseck, J. Z.; Brooks, C. L., III. Approaching protein design with multisite λ dynamics: accurate and scalable mutational folding free energies in T4 lysozyme. *Protein Sci.* **2018**, *27*, 1910−1922.
(178) Hayes, R. L.; Vilseck, J. Z.; Brooks, C. L., III. Addressing intersite coupling unlocks large combinatorial chemical spaces for alchemical free energy methods. *J. Chem. Theory Comp.* **2022**, *18*, 2114−2123.
(179) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. K. All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. *J. Chem. Theory Comp.* **2016**, *12*, 5411−5421.
(180) Ding, X.; Vilseck, J. Z.; Hayes, R. L.; Brooks, C. L., III. Gibbs sampler-based λ-dynamics and Rao-Blackwell estimator for alchemical free energy calculation. *J. Chem. Theory Comp.* **2017**, *13*, 2501−2510.
(181) Vilseck, J. Z.; Ding, X.; Hayes, R. L.; Brooks, C. L., III. Generalizing the discrete Gibbs sampler-based λ-dynamics approach for multisite sampling of many ligands. *J. Chem. Theory Comp.* **2021**, *17*, 3895−3907.
(182) Robo, M. T.; Hayes, R. L.; Ding, X.; Pulawski, B.; Vilseck, J. Z. Fast free energy estimates from λ-dynamics with bias-updated Gibbs sampling. *Nat. Commun.* **2023**, *14*, 8515.
(183) Liu, S.; Wang, L.; Mobley, D. L. Is ring breaking feasible in relative binding free energy calculations? *J. Chem. Inf. and Mod.* **2015**, *55*, 727−735.
(184) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate modeling of scaffold hopping transformations in drug discovery. *J. Chem. Theory Comp.* **2017**, *13*, 42−54.
(185) Hayes, R. L.; Brooks, C. L., III. A strategy for proline and glycine mutations to proteins with alchemical free energy calculations. *J. Comput. Chem.* **2021**, *42*, 1088−1094.
(186) Vilseck, J. Z.; Sohail, N.; Hayes, R. L.; Brooks, C. L., III. Overcoming challenging substituent perturbations with multisite λ-dynamics: a case study targeting β-Secretase 1. *J. Phys. Chem. Lett.* **2019**, *10*, 4875−4880.
(187) Deng, Y. Q.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Chem. Theory Comp.* **2009**, *113*, 2234−2246.
(188) Jiang, W.; Hodoscek, M.; Roux, B. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J. Chem. Theory Comp.* **2009**, *5*, 2583−2588.
(189) Jiang, W.; Roux, B. Free energy perturbation hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comp.* **2010**, *6*, 2559−2565.
(190) Lin, Y. L.; Meng, Y.; Jiang, W.; Roux, B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1664−1669.
(191) Lin, Y. L.; Roux, B. Computational analysis of the binding specificity of Gleevec to Abl, c-Kit, Lck, and c-Src tyrosine kinases. *J. Am. Chem. Soc.* **2013**, *135*, 14741−14753.
(192) Lin, Y. L.; Meng, Y.; Huang, L.; Roux, B. Computational study of Gleevec and G6G reveals molecular determinants of kinase inhibitor selectivity. *J. Am. Chem. Soc.* **2014**, *136*, 14753−14762.
(193) Jiang, W.; Luo, Y.; Maragliano, L.; Roux, B. Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer. *J. Chem. Theory Comp.* **2012**, *8*, 4672−4680.
(194) Deng, Y. Q.; Roux, B. Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **2008**, *128*, 115103.
(195) Im, W.; Bernèche, S.; Roux, B. Generalized solvent boundary potentials for computer simulations. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(196) Ge, X.; Roux, B. Calculation of the standard binding free energy of sparsomycin to the ribosomal peptidyl-transferase P-site using molecular dynamics simulations with restraining potentials. *J. Mol. Recognit.* **2010**, *23*, 128−141.
(197) Ge, X.; Roux, B. Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome. *J. Phys. Chem. B* **2010**, *114*, 9525−9539.
(198) Shivakumar, D.; Deng, Y.; Roux, B. Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theory Comp.* **2009**, *5*, 919−930.
(199) Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J. Reproducibility of free energy calculations across different molecular simulation software packages. *J. Chem. Theory Comp.* **2018**, *14*, 5567−5582.
(200) Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6825−6830.
(201) Gan, W.; Roux, B. Binding specificity of SH2 domains: insight from free energy simulations. *Proteins* **2009**, *74*, 996−1007.
(202) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Bolztmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(203) Nandigrami, P.; Szczepaniak, F.; Boughter, C. T.; Dehez, F.; Chipot, C.; Roux, B. Computational assessment of protein-protein binding specificity within a family of synaptic surface receptors. *J. Phys. Chem. B* **2022**, *126*, 7510−7527.
(204) Chen, Y.; Roux, B. Efficient hybrid non-equilibrium molecular dynamics - Monte Carlo simulations with symmetric momentum reversal. *J. Chem. Phys.* **2014**, *141*, 114107.
(205) Chen, Y.; Roux, B. Generalized metropolis acceptance criterion for hybrid non-equilibrium molecular dynamics - Monte Carlo simulations. *J. Chem. Phys.* **2015**, *142*, 024101.
(206) Chen, Y.; Roux, B. Enhanced sampling of an atomic model with hybrid nonequilibrium molecular dynamics: Monte Carlo simulations guided by a coarse-grained model. *J. Chem. Theory Comp.* **2015**, *11*, 3572−3583.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(207) Chen, Y.; Roux, B. Constant-pH hybrid nonequilibrium molecular dynamics-Monte Carlo simulation method. *J. Chem. Theory Comp.* **2015**, *11*, 3919−3931.
(208) Suh, D.; Jo, S.; Jiang, W.; Chipot, C.; Roux, B. String method for protein-protein binding free-energy calculations. *J. Chem. Theory Comp.* **2019**, *15*, 5829−5844.
(209) Chen, H.; Chipot, C. Enhancing sampling with free-energy calculations. *Curr. Op. Struct. Biol.* **2022**, *77*, 102497.
(210) Liu, X.; Brooks, C. L., III. Enhanced sampling of buried charges in free energy calculations using replica exchange with charge tempering. *J. Chem. Theory Comp.* **2024**, *20*, 1051−1061.
(211) Lee, J.; Miller, B. T.; Brooks, B. R. Computational scheme for pH-dependent binding free energy calculation with explicit solvent. *Protein Sci.* **2016**, *25*, 231−243.
(212) Allen, T. W.; Andersen, O. S.; Roux, B. Ion permeation through a narrow channel: using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields. *Biophys. J.* **2006**, *90*, 3447−3468.
(213) Wieder, M.; Fleck, M.; Braunsfeld, B.; Boresch, S. Alchemical free energy simulations without speed limits. A generic framework to calculate free energy differences independent of the underlying molecular dynamics program. *J. Comput. Chem.* **2022**, *43*, 1151−1160.
(214) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. II. improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(215) Reinhardt, M.; Grubmüller, H. Determining free-energy differences through variationally derived intermediates. *J. Chem. Theory Comp.* **2020**, *16*, 3504−3512.
(216) Konig, G.; Brooks, B. R.; Thiel, W.; York, D. M. On the convergence of multi-scale free energy simulations. *Mol. Sim.* **2018**, *44*, 1062−1081.
(217) Blondel, A. Ensemble variance in free energy calculations by thermodynamic integration: theory, optimal "alchemical" path, and practical solutions. *J. Comput. Chem.* **2004**, *25*, 985−993.
(218) Pham, T. T.; Shirts, M. R. Optimal pairwise and non-pairwise alchemical pathways for free energy calculations of molecular transformation in solution phase. *J. Chem. Phys.* **2012**, *136*, 124120.
(219) König, G.; Ries, B.; Hünenberger, P. H.; Riniker, S. Efficient alchemical intermediate states in free energy calculations using λ-enveloping distribution sampling. *J. Chem. Theory Comp.* **2021**, *17*, 5805−5815.
(220) König, G.; Glaser, N.; Schroeder, B.; KubincovÃ, A.; Hünenberger, P. H.; Riniker, S. An alternative to conventional λ-intermediate states in alchemical free energy calculations: λ-enveloping distribution sampling. *J. Chem. Inf. and Mod.* **2020**, *60*, 5407−5423.
(221) Narayana, N.; Matthews, D. A.; Howell, E. E.; Xuong, N.-h. A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. *Nat. Struct. Biol.* **1995**, *2*, 1018−1025.
(222) Dam, J.; Rose, T.; Goldberg, M. E.; Blondel, A. Complementation between dimeric mutants as a probe of dimer-dimer interactions in tetrameric dihydrofolate reductase encoded by R67 plasmid of E. coli. *J. Mol. Biol.* **2000**, *302*, 235−250.
(223) Dam, J.; Blondel, A. Effect of multiple symmetries on the association of R67 DHFR subunits bearing interfacial complementing mutations. *Protein Sci.* **2004**, *13*, 1−14.
(224) Straatsma, T. P.; Berendsen, H. J. C.; Stam, A. J. Estimation of statistical errors in molecular simulation calculations. *Mol. Phys.* **1986**, *57*, 89−95.
(225) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. Constant-pH molecular dynamics using continuous titration coordinates. *PROTEINS: Struct., Func. and Bioinf.* **2004**, *56*, 738−752.
(226) Khandogin, J.; Brooks, C. L., III. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* **2005**, *89*, 141−157.
(227) Wallace, J. A.; Shen, J. K. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comp.* **2011**, *7*, 2617−2629.
(228) Huang, Y.; Chen, W.; Dotson, D. L.; Beckstein, O.; Shen, J. Mechanism of pH-dependent activation of the sodium-proton antiporter NhaA. *Nat. Commun.* **2016**, *7*, 12940.
(229) Im, W.; Feig, M.; Brooks, C. L., III. An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **2003**, *85*, 2900−2918.
(230) Wallace, J. A.; Shen, J. K. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* **2012**, *137*, 184105.
(231) Chen, W.; Wallace, J. A.; Yue, Z.; Shen, J. K. Introducing titratable water to all-atom molecular dynamics at constant pH. *Biophys. J.* **2013**, *105*, L15−L17.
(232) Morrow, B. H.; Payne, G. F.; Shen, J. pH-responsive self-assembly of polysaccharide through a rugged energy landscape. *J. Am. Chem. Soc.* **2015**, *137*, 13024−13030.
(233) Huang, Y.; Yue, Z.; Tsai, C.-C.; Henderson, J. A.; Shen, J. Predicting catalytic proton donors and nucleophiles in enzymes: how adding dynamics helps elucidate the structure-function relationships. *J. Phys. Chem. Lett.* **2018**, *9*, 1179−1184.
(234) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.
(235) Bürgi, R.; Kollman, P. A.; Van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *PROTEINS: Struct., Func. and Bioinf.* **2002**, *47*, 469−480.
(236) Dlugosz, M.; Antosiewicz, J. M. Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chem. Phys.* **2004**, *302*, 161−170.
(237) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.
(238) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *PROTEINS: Struct., Func. and Bioinf.* **2011**, *79*, 3420−3436.
(239) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J. Chem. Theory Comp.* **2014**, *10*, 1341−1352.
(240) Stern, H. A. Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.* **2007**, *126*, 164112.
(241) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-pH molecular dynamics simulations for large biomolecular systems. *J. Chem. Theory Comp.* **2017**, *13*, 5933−5944.
(242) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comp.* **2011**, *7*, 1962−1978.
(243) Damjanović, A.; Miller, B. T.; Okur, A.; Brooks, B. R. Reservoir pH replica exchange. *J. Chem. Phys.* **2018**, *149*, 072321.
(244) Damjanovic, A.; Chen, A. Y.; Rosenberg, R. L.; Roe, D. R.; Wu, X.; Brooks, B. R. Protonation state of the selectivity filter of bacterial voltage-gated sodium channels is modulated by ions. *PROTEINS: Struct., Func. and Bioinf.* **2020**, *88*, 527−539.
(245) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: p K a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.
(246) Ghosh, N.; Cui, Q. PKa of Residue 66 in Staphylococal nuclease. I. Insights from QM/MM Simulations with Conventional Sampling. *J. Phys. Chem. B* **2008**, *112*, 8387−8397.
(247) Chen, A. Y.; Brooks, B. R.; Damjanovic, A. Ion channel selectivity through ion-modulated changes of selectivity filter PKa values. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, No. e2220343120.
(248) Lazaridis, T.; Hummer, G. Classical molecular dynamics with mobile protons. *J. Chem. Inf. and Mod.* **2017**, *57*, 2833−2845.
(249) Lazaridis, T.; Sepehri, A. Amino acid deprotonation rates from classical force fields. *J. Chem. Phys.* **2022**, *157*, 085101.
(250) Lazaridis, T. Molecular origins of asymmetric proton conduction in the influenza M2 channel. *Biophys. J.* **2023**, *122*, 90−98.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(251) Lazaridis, T. Proton paths in models of the Hv1 proton channel. *J. Phys. Chem. B* **2023**, *127*, 7937−7945.
(252) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(253) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(254) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta - General Subjects* **2015**, *1850*, 872−877.
(255) Paschek, D.; García, A. E. Reversible temperature and pressure denaturation of a protein fragment: a replica exchange molecular dynamics simulation study. *Phys. Rev. Lett.* **2004**, *93*, 10−13.
(256) Wu, X.; Hodoscek, M.; Brooks, B. R. Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling. *J. Chem. Phys.* **2012**, *137*, 1−13.
(257) Okur, A.; Miller, B. T.; Joo, K.; Lee, J.; Brooks, B. R. Generating reservoir conformations for replica exchange through the use of the conformational space annealing method. *J. Chem. Theory Comp.* **2013**, *9*, 1115−1124.
(258) Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C. Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J. Chem. Theory Comp.* **2007**, *3*, 557−568.
(259) Roitberg, A. E.; Okur, A.; Simmerling, C. Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. *J. Phys. Chem. B* **2007**, *111*, 2415−2418.
(260) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(261) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E.; Cheatham, T. E., III. Multi-dimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comp.* **2014**, *10*, 492−499.
(262) Woodcock III, H. L.; Miller, B. T.; Hodoscek, M.; Okur, A.; Larkin, J. D.; Ponder, J. W.; Brooks, B. R. MSCALE: a general utility for multiscale modeling. *J. Chem. Theory Comp.* **2011**, *7*, 1208−1219.
(263) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. Targeted molecular-dynamics simulation of conformational change - application to the T ↔ R transition in insulin. *Mol. Sim.* **1993**, *10*, 291−308.
(264) Apostolakis, J.; Ferrara, P.; Caflisch, A. Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water. *J. Chem. Phys.* **1999**, *110*, 2099−2108.
(265) van der Vaart, A.; Karplus, M. Simulation of conformational transitions by the restricted perturbation-targeted molecular dynamics method. *J. Chem. Phys.* **2005**, *122*, 114903.
(266) van der Vaart, A.; Karplus, M. Minimum free energy pathways and free energy profiles for conformational transitions based on atomistic molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 164106.
(267) Ovchinnikov, V.; Karplus, M. Analysis and elimination of a bias in targeted molecular dynamics simulations of conformational transitions: application to calmodulin. *J. Phys. Chem. B* **2012**, *116*, 8584−8603.
(268) Hooft, R.; Vaneijk, B.; Kroon, J. An adaptive umbrella sampling procedure in conformational-analysis using molecular-dynamics and its application to glycol. *J. Chem. Phys.* **1992**, *97*, 6690−6694.
(269) Bartels, C.; Karplus, M. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J. Comput. Chem.* **1997**, *18*, 1450−1462.
(270) Ovchinnikov, V.; Cecchini, M.; Karplus, M. A simplified confinement method for calculating absolute free energies and free energy and entropy differences. *J. Phys. Chem. B* **2013**, *117*, 750−762.
(271) Esque, J.; Cecchini, M. Accurate calculation of conformational free energy differences in explicit water: the confinement-solvation free energy approach. *J. Phys. Chem. B* **2015**, *119*, 5194−5207.
(272) Villemot, F.; Peguero-Tejada, A.; van der Vaart, A. Calculation of conformational free energies by confinement simulations in explicit water with implicit desolvation. *Mol. Sim.* **2018**, *44*, 1082−1089.
(273) van der Vaart, A.; Orndorff, P. B.; Le Phan, S. T. Calculation of conformational free energies with the focused confinement method. *J. Chem. Theory Comp.* **2019**, *15*, 6760−6768.
(274) Orndorff, P. B.; Le Phan, S. T.; Li, K. H.; van der Vaart, A. Conformational free-energy differences of large solvated systems with the focused confinement method. *J. Chem. Theory Comp.* **2020**, *16*, 5163−5173.
(275) Maragakis, P.; van der Vaart, A.; Karplus, M. Gaussian-mixture umbrella sampling. *J. Phys. Chem. B* **2009**, *113*, 4664−4673.
(276) Spiriti, J.; Binder, J. K.; Levitus, M.; van der Vaart, A. Cy3-DNA stacking interactions strongly depend on the identity of the terminal basepair. *Biophys. J.* **2011**, *100*, 1049−1057.
(277) Andricioaei, I.; Straub, J. E. On Monte Carlo and molecular dynamics methods inspired by Tsallis statistics: methodology, optimization, and application to atomic clusters. *J. Chem. Phys.* **1997**, *107*, 9117−9124.
(278) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13749−13754.
(279) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(280) Kamberaj, H.; van der Vaart, A. Multiple scaling replica exchange for the conformational sampling of biomolecules in explicit water. *J. Chem. Phys.* **2007**, *127*, 234102.
(281) Gray, G. M.; Thiessen, B.; van der Vaart, A. Secondary structure of peptides mimicking the Gly-rich regions of major ampullate spidroin protein 1 and 2. *Biophys. Chem.* **2022**, *284*, 106783.
(282) Elber, R.; Karplus, M. A method for determining reaction paths in large biomolecules: application to myoglobin. *Chem. Phys. Lett.* **1987**, *139*, 375−380.
(283) Jónsson, G. M.; Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B., Ciccotti, G., Coker, D., Eds.; World Scientific: Singapore, 1998; pp 385−404.
(284) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(285) Ren, W.; Vanden-Eijnden, E.; Maragakis, P.; E, W. Transition pathways in complex systems: application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* **2005**, *123*, 134109.
(286) Huo, S.; Straub, J. E. The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature. *J. Chem. Phys.* **1997**, *107*, 5000−5006.
(287) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From a to b in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
(288) E, W.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.
(289) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
(290) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, 194101.
(291) Gardiner, C. W. *Handbook of stochastic methods*, 3rd ed.; Springer-Verlag: Berlin, 2003.
(292) E, W.; Vanden-Eijnden, E. In *Multiscale Modeling and Simulation*; Attinger, S., Koumoutsakos, P., Eds.; Springer-Verlag: Berlin, 2004.
(293) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(294) E, W.; Ren, W.; Vanden-Eijnden, E. Transition pathways in complex systems: reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* **2005**, *413*, 242−247.
(295) E, W.; Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.
(296) Ovchinnikov, V.; Karplus, M. Investigations of α-helix ↔ β-sheet transition pathways in a miniprotein using the finite-temperature string method. *J. Chem. Phys.* **2014**, *140*, 175103.
(297) Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q. QM/MM free energy simulations: recent progress and challenges. *Mol. Sim.* **2016**, *42*, 1056−1078.
(298) Pan, A.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.
(299) Maragliano, L.; Roux, B.; Vanden-Eijnden, E. Comparison between mean forces and swarms-of-trajectories string methods. *J. Chem. Theory Comp.* **2014**, *10*, 524−533.
(300) Roux, B. String method with swarms-of-trajectories, mean drifts, lag time, and committor. *J. Phys. Chem. A* **2021**, *125*, 7558−7571.
(301) Gan, W.; Yang, S.; Roux, B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys. J.* **2009**, *97*, L8−L10.
(302) Meng, Y.; Lin, Y.-l.; Roux, B. Computational study of the "DFG-flip" conformational transition in c-Abl and c-Src tyrosine kinases. *J. Phys. Chem. B* **2015**, *119*, 1443−1456.
(303) Harder, E.; Walters, D. E.; Bodnar, Y. D.; Faibish, R. S.; Roux, B. Molecular dynamics study of a polymeric reverse osmosis membrane. *J. Phys. Chem. B* **2009**, *113*, 10177−10182.
(304) Luo, Y.; Harder, E.; Faibish, R. S.; Roux, B. Computer simulations of water flux and salt permeability of the reverse osmosis FT-30 aromatic polyamide membrane. *J. Membr. Sci.* **2011**, *384*, 1−9.
(305) Dickson, B. M.; Huang, H.; Post, C. B. Unrestrained computation of free energy along a path. *J. Phys. Chem. B* **2012**, *116*, 11046−11055.
(306) Dickson, B. M. Approaching a parameter-free metadynamics. *Phys. Rev. E* **2011**, *84*, 037701.
(307) E, W.; Ren, W.; Vanden-Eijnden, E. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(308) Wu, H.; Post, C. B. Protein conformational transitions from all-atom adaptively biased path optimization. *J. Chem. Theory Comp.* **2018**, *14*, 5372−5382.
(309) Wu, H.; Huang, H.; Post, C. B. All-atom adaptively biased path optimization of Src kinase conformational inactivation: switched electrostatic network in the concerted motion of αC helix and the activation loop. *J. Chem. Phys.* **2020**, *153*, 175101.
(310) Lee Woodcock, H.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F., III; Brooks, B. R. Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.* **2003**, *109*, 140−148.
(311) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978−9985.
(312) Chu, J.-W.; Trout, B. L.; Brooks, B. R. A super-linear minimization scheme for the nudged elastic band method. *J. Chem. Phys.* **2003**, *119*, 12708.
(313) Brokaw, J. B.; Haas, K. R.; Chu, J.-W. Reaction path optimization with holonomic constraints and kinetic energy potentials. *J. Chem. Theory Comp.* **2009**, *5*, 2050−2061.
(314) Haas, K.; Chu, J.-W. Decomposition of energy and free energy changes by following the flow of work along reaction path. *J. Chem. Phys.* **2009**, *131*, 144105.
(315) Cho, H. M.; Gross, A. S.; Chu, J.-W. Dissecting force interactions in cellulose deconstruction reveals the required solvent versatility for overcoming biomass recalcitrance. *J. Am. Chem. Soc.* **2011**, *133*, 14033−14041.
(316) Lin, Y.; Beckham, G. T.; Himmel, M. E.; Crowley, M. F.; Chu, J.-W. Endoglucanase peripheral loops facilitate complexation of glucan chains on cellulose via adaptive coupling to the emergent substrate structures. *J. Phys. Chem. B* **2013**, *117*, 10750−10758.
(317) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic analysis of large systems. I. methodology. *J. Comput. Chem.* **1995**, *16*, 1522−1542.
(318) Chen, Y.-T.; Yang, H.; Chu, J.-W. Structure-mechanics statistical learning unravels the linkage between local rigidity and global flexibility in nucleic acids. *Chem. Sci.* **2020**, *11*, 4969−4979.
(319) Raj, N.; Click, T.; Yang, H.; Chu, J.-W. Mechanical couplings of protein backbone and side chains exhibit scale-free network properties and specific hotspots for function. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5309−5320.
(320) Raj, N.; Click, T. H.; Yang, H.; Chu, J.-W. Structure-mechanics statistical learning uncovers mechanical relay in proteins. *Chem. Sci.* **2022**, *13*, 3688−3696.
(321) Chen, Y.-T.; Yang, H.; Chu, J.-W. Mechanical codes of chemical-scale specificity in DNA motifs. *Chem. Sci.* **2023**, *14*, 10155−10166.
(322) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed molecular dynamics: a simple and general technique for accelerating rare event kinetics and mapping free energy in large molecular systems. *J. Phys. Chem. B* **2009**, *113*, 16603−16611.
(323) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed molecular dynamics: decorrelation time scales and the kinetic master equation. *J. Chem. Theory Comp.* **2011**, *7*, 1244−1252.
(324) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(325) O'Connor, M.; Paci, E.; McIntosh-Smith, S.; Glowacki, D. R. Adaptive free energy sampling in multidimensional collective variable space using boxed molecular dynamics. *Faraday Discuss* **2016**, *195*, 395−419.
(326) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(327) Darve, E.; Wilson, M. A.; Pohorille, A. Calculating free energies using a scaled-force molecular dynamics algorithm. *Molecular Simulation* **2002**, *28*, 113−144.
(328) Hénin, J.; Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **2004**, *121*, 2904−2914.
(329) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(330) Fu, H.; Shao, X.; Chipot, C.; Cai, W. Extended adaptive biasing force algorithm. an on-the-fly implementation for accurate free-energy calculations. *J. Chem. Theory Comp.* **2016**, *12*, 3506−3513.
(331) Lesage, A.; Leliévre, T.; Stoltz, G.; Hénin, J. Smoothed biasing forces yield unbiased free energies with the extended-system adaptive biasing force method. *J. Phys. Chem. B* **2017**, *121*, 3676−3685.
(332) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *128*, 144120.
(333) Leliévre, T.; Rousset, M.; Stoltz, G. *Free Energy Computations*; Imperial College Press: 2010.
(334) Leliévre, T.; Rousset, M.; Stoltz, G. Computation of free energy profiles with parallel adaptive dynamics. *J. Chem. Phys.* **2007**, *126*, 134111.
(335) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20227−20232.
(336) Zheng, L.; Yang, W. Practically efficient and robust free energy calculations: double-integration orthogonal space tempering. *J. Chem. Theory Comp.* **2012**, *8*, 810−823.
(337) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comp.* **2010**, *6*, 35−47.
(338) Stone, A. J. *The theory of intermolecular forces*; Clarendon Press: Oxford, U.K., 1996; Vol. *32*.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(339) Stone, A. J. Distributed multipole analysis: stability for large basis sets. *J. Chem. Theory Comp.* **2005**, *1*, 1128−1132.
(340) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(341) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A* **2003**, *107*, 10353−10359.
(342) Bereau, T.; Kramer, C.; Meuwly, M. Leveraging symmetries of static atomic multipole electrostatics in molecular dynamics simulations. *J. Chem. Theory Comp.* **2013**, *9*, 5450−5459.
(343) Cardamone, S.; Popelier, P. L. A. Prediction of conformationally dependent atomic multipole moments in carbohydrates. *J. Comput. Chem.* **2015**, *36*, 2361−2373.
(344) Rein, R. On physical properties and interactions of polyatomic molecules with application to molecular recognition in biology. *Adv. Quantum Chem.* **1973**, *7*, 335−396.
(345) Maroulis, G. Electric polarizability and hyperpolarizability of carbon monoxide. *J. Phys. Chem.* **1996**, *100*, 13466−13473.
(346) Roco, J. M. M.; Calvo Hernández, A.; Velasco, S. Far-infrared permanent and induced dipole absorption of diatomic molecules in rare-gas fluids. I. spectral theory. *J. Chem. Phys.* **1995**, *103*, 9161−9174.
(347) Roco, J. M. M.; Medina, A.; Calvo Hernández, A.; Velasco, S. Far-infrared permanent and induced dipole absorption of diatomic molecules in rare-gas fluids. II. application to the CO−Ar system. *J. Chem. Phys.* **1995**, *103*, 9175−9186.
(348) Straub, J. E.; Karplus, M. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: ligand dynamics in the 1st 10 ps. *Chem. Phys.* **1991**, *158*, 221−248.
(349) Nutt, D.; Meuwly, M. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy Myoglobin. *Biophys. J.* **2003**, *85*, 3612−3623.
(350) Nutt, D.; Meuwly, M. CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein function. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5998−6002.
(351) Plattner, N.; Meuwly, M. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.* **2008**, *94*, 2505−2515.
(352) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the σ − hole. *J. Mol. Model.* **2007**, *13*, 291−296.
(353) Murray, J. S.; Lane, P.; Politzer, P. Expansion of the σ − hole concept. *J. Mol. Model.* **2009**, *15*, 723−729.
(354) Politzer, P.; Murray, J. S.; Clark, T. Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748−7757.
(355) Jackson, J. D. *Classical electrodynamics*; John Wiley & Sons: New York, 1998.
(356) Devereux, M.; Raghunathan, S.; Fedorov, D. G.; Meuwly, M. A novel, computationally efficient multipolar model employing distributed charges for molecular dynamics simulations. *J. Chem. Theory Comp.* **2014**, *10*, 4229−4241.
(357) Unke, O. T.; Devereux, M.; Meuwly, M. Minimal distributed charges: Multipolar quality at the cost of point charge electrostatics. *J. Chem. Phys.* **2017**, *147*, 161712.
(358) Boittier, E. D.; Devereux, M.; Meuwly, M. Molecular dynamics with conformationally dependent, distributed charges. *J. Chem. Theory Comp.* **2022**, *18*, 7544−7554.
(359) Boittier, E. D.; Devereux, M.; Töpfer, K.; Meuwly, M. Kernel-based minimally distributed charges: a conformationally dependent ESP-model for molecular simulations. *J. Chem. Theory Comp.* **2024**, in press. DOI: 10.48550/arXiv.2406.00513.
(360) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chem. Rev.* **2021**, *121*, 10142−10186.
(361) Manzhos, S.; Carrington, T., Jr Neural network potential energy surfaces for small molecules and reactions. *Chem. Rev.* **2021**, *121*, 10187−10217.
(362) Meuwly, M. Machine learning for chemical reactions. *Chem. Rev.* **2021**, *121*, 10218−10239.
(363) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian process regression for materials and molecules. *Chem. Rev.* **2021**, *121*, 10073−10141.
(364) Houston, P. L.; Qu, C.; Yu, Q.; Conte, R.; Nandi, A.; Li, J. K.; Bowman, J. M. PESPIP: software to fit complex molecular and many-body potential energy surfaces with permutationally invariant polynomials. *J. Chem. Phys.* **2023**, *158*, 044109.
(365) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Intern. Rev. Phys. Chem.* **2009**, *28*, 577−606.
(366) Schuett, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Mueller, K. R. SchNet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.
(367) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; E, W. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. *Advances in Neural Information Processing Systems*. **2018**.
(368) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, No. e1603015.
(369) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
(370) Ho, T.-S.; Rabitz, H. A general method for constructing multidimensional molecular potential energy surfaces from ab initio calculations. *J. Chem. Phys.* **1996**, *104*, 2584−2597.
(371) Unke, O. T.; Meuwly, M. Toolkit for the construction of reproducing kernel-based representations of data: application to multidimensional potential energy surfaces. *J. Chem. Inf. and Mod.* **2017**, *57*, 1923−1931.
(372) Meuwly, M.; Hutson, J. M. The potential energy surface and near-dissociation states of He-H_2^+. *J. Chem. Phys.* **1999**, *110*, 3418−3427.
(373) Koner, D.; Bemish, R. J.; Meuwly, M. The $C(^3P)$ + $NO(X^2\Pi)$ → $O(^3P)$ + $CN(X^2\Sigma^+)$, $N(^2D)/N(^4S)$ + $CO(X^1\Sigma^+)$ reaction: rates, branching ratios, and final states from 15 to 20000 K. *J. Chem. Phys.* **2018**, *149*, 094305.
(374) San Vicente Veliz, J. C.; Koner, D.; Schwilk, M.; Bemish, R. J.; Meuwly, M. The $N(^4S)$ + $O_2(X^3\Sigma)$ ↔ $O(^3P)$ + $NO(X^2\Pi)$ reaction: thermal and vibrational relaxation rates for the $^2A'$, $^4A'$ and $^2A''$ states. *Phys. Chem. Chem. Phys.* **2020**, *22*, 3927−3939.
(375) Koner, D.; Meuwly, M. Permutationally invariant, reproducing kernel-based potential energy surfaces for polyatomic molecules: from formaldehyde to acetone. *J. Chem. Theory Comp.* **2020**, *16*, 5474−5484.
(376) Song, K.; Käser, S.; Töpfer, K.; Vazquez-Salazar, L. I.; Meuwly, M. PhysNet meets CHARMM: A framework for routine machine learning/molecular mechanics simulations. *J. Chem. Phys.* **2023**, *159*, 024125.
(377) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM simulations: what are some of the "burning issues". *J. Phys. Chem. B* **2021**, *125*, 689−702.
(378) Gastegger, M.; Schütt, K. T.; Müller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.* **2021**, *12*, 11473−11483.
(379) Töpfer, K.; Käser, S.; Meuwly, M. Double proton transfer in hydrated formic acid dimer: interplay of spatial symmetry and solvent-generated force on reactivity. *Phys. Chem. Chem. Phys.* **2022**, *24*, 13869−13882.
(380) Lier, B.; Poliak, P.; Marquetand, P.; Westermayr, J.; Oostenbrink, C. BuRNN: buffer region neural network approach for polarizable-embedding neural network/molecular mechanics simulations. *J. Phys. Chem. Lett.* **2022**, *13*, 3812−3818.

(381) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(382) Upadhyay, M.; Meuwly, M. Thermal and vibrationally activated decomposition of the Syn-CH_3CHOO Criegee intermediate. *ACS Earth Space Chem* **2021**, *5*, 3396−3406.
(383) Hänninen, V.; Murdachaew, G.; Nathanson, G. M.; Gerber, R. B.; Halonen, L. Ab initio molecular dynamics studies of formic acid dimer colliding with liquid water. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23717−23725.
(384) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K. R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
(385) Käser, S.; Richardson, J. O.; Meuwly, M. Transfer learning for affordable and high-quality tunneling splittings from instanton calculations. *J. Chem. Theory Comp.* **2022**, *18*, 6840−6850.
(386) Käser, S.; Meuwly, M. Transfer learned potential energy surfaces: accurate anharmonic vibrational dynamics and dissociation energies for the formic acid monomer and dimer. *Phys. Chem. Chem. Phys.* **2022**, *24*, 5269−5281.
(387) Kidwell, N. M.; Li, H.; Wang, X.; Bowman, J. M.; Lester, M. I. Unimolecular dissociation dynamics of vibrationally activated CH_3CHOO Criegee intermediates to OH radical products. *Nat. Chem.* **2016**, *8*, 509−514.
(388) Käser, S.; Unke, O. T.; Meuwly, M. Isomerization and decomposition reactions of acetaldehyde relevant to atmospheric processes from dynamics simulations on neural network-based potential energy surfaces. *J. Chem. Phys.* **2020**, *152*, 214304.
(389) Upadhyay, M.; Töpfer, K.; Meuwly, M. Molecular simulation for atmospheric reactions: non-equilibrium dynamics, roaming, and glycolaldehyde formation following photoinduced decomposition of syn-acetaldehyde oxide. *J. Phys. Chem. Lett.* **2024**, *15*, 90−96.
(390) Tan, M.-L.; Tran, K. N.; Pickard IV, F. C.; Simmonett, A. C.; Brooks, B. R.; Ichiye, T. Molecular multipole potential energy functions for water. *J. Phys. Chem. B* **2016**, *120*, 1833−1842.
(391) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(392) Huang, J.; Simmonett, A. C.; Pickard, F. C.; MacKerell, A. D., Jr.; Brooks, B. R. Mapping the Drude polarizable force field onto a multipole and induced dipole model. *J. Chem. Phys.* **2017**, *147*, 161702.
(393) Simmonett, A. C.; Pickard, F. C.; Shao, Y.; Cheatham, T. E.; Brooks, B. R. Efficient treatment of induced dipoles. *J. Chem. Phys.* **2015**, *143*, 074115.
(394) Simmonett, A. C.; Pickard, F. C.; Ponder, J. W.; Brooks, B. R. An empirical extrapolation scheme for efficient treatment of induced dipoles. *J. Chem. Phys.* **2016**, *145*, 164101.
(395) Gao, J.; Habibollazadeh, D.; Shao, L. A polarizable intermolecular potential function for simulation of liquid alcohols. *J. Phys. Chem.* **1995**, *99*, 16460−16467.
(396) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. Simulation of liquid amides using a polarizable intermolecular potential function. *J. Phys. Chem.* **1996**, *100*, 2689−2697.
(397) Xie, W.; Pu, J.; MacKerell, A. D., Jr.; Gao, J. Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *J. Chem. Theory Comp.* **2007**, *3*, 1878−1889.
(398) Xie, W.; Pu, J.; Gao, J. A coupled polarization-matrix inversion and iteration approach for accelerating the dipole convergence in a polarizable potential function. *J. Phys. Chem. A* **2009**, *113*, 2109−2116.
(399) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504−1514.
(400) Patel, S.; Brooks, C. L., III. CHARMM fluctuating charge force field for proteins: I parametrization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1−16.
(401) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. A simple polarizable model of water based on classical Drude oscillator. *J. Chem. Phys.* **2003**, *119*, 5185−5197.
(402) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(403) Angyan, J. G.; Colonna-Cesari, F.; Tapia, O. Analytical first and second energy derivatives in the polarization model. *Chem. Phys. Lett.* **1990**, *166*, 180−188.
(404) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.
(405) van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.
(406) Vesely, F. J. N-particle dynamics of polarizable Stockmayer-type molecules. *J. Chem. Phys.* **1977**, *24*, 361−371.
(407) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. Solvation free energies of small amides and amines from molecular dynamics/free energy perturbation simulations using pairwise additive and many-body polarizable potentials. *J. Phys. Chem.* **1995**, *99*, 11575−11583.
(408) Van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J. Molecular dynamics simulation of polarizable water by an extended Lagrangian method. *Mol. Phys.* **1992**, *77*, 239−255.
(409) Hoover, W. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(410) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(411) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(412) Winkler, R. G.; Kraus, V.; Reineker, P. Time reversible and phase-space conserving molecular dynamics at constant temperature. *J. Chem. Phys.* **1995**, *102*, 9018−9025.
(413) Ahlstrom, P.; Wallqvist, A.; Engstrom, S.; Jonsson, B. A molecular dynamics study of polarizable water. *Mol. Phys.* **2006**, *68*, 563−581.
(414) Applequist, J.; Carl, J. R.; Fung, K.-K. An atom dipole interaction model for molecular polarizability. application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.
(415) Kim, B.; Shao, Y.; Pu, J. Doubly polarized QM/MM with machine learning chaperone polarizability. *J. Chem. Theory Comp.* **2021**, *17*, 7682−7695.
(416) Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. Effect of electrostatic force truncation on interfacial and transport properties of water. *J. Phys. Chem.* **1996**, *100*, 17011−17020.
(417) Klauda, J. B.; Wu, X. W.; Pastor, R. W.; Brooks, B. R. Long-range Lennard-Jones and electrostatic interactions in interfaces: application of the isotropic periodic sum method. *J. Phys. Chem. B* **2007**, *111*, 4393−4400.
(418) Venable, R. M.; Chen, L. E.; Pastor, R. W. Comparison of the extended isotropic periodic sum and particle mesh Ewald methods for simulations of lipid bilayers and monolayers. *J. Phys. Chem. B* **2009**, *113*, 5855−5862.
(419) Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *J. Chem. Theory Comp.* **2013**, *9*, 3527−3537.
(420) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830−7843.
(421) Wennberg, C. L.; Murtola, T.; Pall, S.; Abraham, M. J.; Hess, B.; Lindahl, E. Direct-space corrections enable fast and accurate Lorentz-Berthelot combination rule Lennard-Jones lattice summation. *J. Chem. Theory Comp.* **2015**, *11*, 5737−5746.
(422) Leonard, A. N.; Simmonett, A. C.; Pickard, F. C.; Huang, J.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Comparison

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

of additive and polarizable models with explicit treatment of long-range Lennard-Jones interactions using alkane simulations. *J. Chem. Theory Comp.* **2018**, *14*, 948−958.
(423) Perram, J. W.; Petersen, H. G.; De Leeuw, S. W. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. *Mol. Phys.* **1988**, *65*, 875−893.
(424) Wu, X. W.; Brooks, B. R. Isotropic periodic sum: a method for the calculation of long-range interactions. *J. Chem. Phys.* **2005**, *122*, 044107.
(425) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426−434.
(426) Attanasio, F.; Convertino, M.; Magno, A.; Caflisch, A.; Corazza, A.; Haridas, H.; Esposito, G.; Cataldo, S.; Pignataro, B.; Milardi, D.; et al. Carnosine inhibits Aβ42 aggregation by perturbing the h-bond network in and around the central hydrophobic cluster. *ChemBioChem* **2013**, *14*, 583−592.
(427) Carballo-Pacheco, M.; Vancea, I.; Strodel, B. Extension of the FACTS implicit solvation model to membranes. *J. Chem. Theory Comp.* **2014**, *10*, 3163−3176.
(428) Zoete, V.; Schuepbach, T.; Bovigny, C.; Chaskar, P.; Daina, A.; Röhrig, U. F.; Michielin, O. Attracting cavities for docking. Replacing the rough energy landscape of the protein by a smooth attracting landscape. *J. Comput. Chem.* **2016**, *37*, 437−447.
(429) Röhrig, U. F.; Goullieux, M.; Bugnon, M.; Zoete, V. Attracting cavities 2.0: improving the flexibility and robustness for small-molecule docking. *J. Chem. Inf. and Mod.* **2023**, *63*, 3925−3940.
(430) Conti, S.; Ovchinnikov, V.; Karplus, M. ppdx: automated modeling of protein−protein interaction descriptors for use with machine learning. *J. Comput. Chem.* **2022**, *43*, 1747−1757.
(431) Lazaridis, T. Effective energy function for proteins in lipid membranes. *PROTEINS: Struct., Func. and Bioinf.* **2003**, *52*, 176−192.
(432) Lazaridis, T. Implicit solvent simulations of peptide interactions with anionic lipid membranes. *PROTEINS: Struct., Func. and Bioinf.* **2005**, *58*, 518−527.
(433) Lazaridis, T.; Karplus, M. Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics* **1999**, *35*, 133−152.
(434) Lazaridis, T. Structural determinants of transmembrane beta-barrels. *J. Chem. Theory Comp.* **2005**, *1*, 716−722.
(435) Mihajlovic, M.; Lazaridis, T. Antimicrobial peptides bind more strongly to membrane pores. *Biochim. et Biophys. Acta Biomembranes* **2010**, *1798*, 1494−1502.
(436) He, Y.; Prieto, L.; Lazaridis, T. Modeling peptide binding to anionic membrane pores. *J. Comput. Chem.* **2013**, *34*, 1463−1475.
(437) Lazaridis, T.; He, Y.; Prieto, L. Membrane interactions and pore formation by the antimicrobial peptide protegrin. *Biophys. J.* **2013**, *104*, 633−642.
(438) Prieto, L.; He, Y.; Lazaridis, T. Protein arcs may form stable pores in lipid membranes. *Biophys. J.* **2014**, *106*, 154−161.
(439) He, Y.; Lazaridis, T. Activity determinants of helical antimicrobial peptides: a large-scale computational study. *PLOS One* **2013**, *8*, No. e66440.
(440) Rahaman, A.; Lazaridis, T. Alamethicin pore formation: a thermodynamic approach. *Biochim. et Biophys. Acta Biomembranes* **2014**, *1838*, 1440−1447.
(441) Lipkin, R. B.; Lazaridis, T. Implicit membrane investigation of the stability of antimicrobial peptide β-barrels and arcs. *J. Membr. Biol.* **2015**, *248*, 469−486.
(442) Lipkin, R.; Lazaridis, T. Computational prediction of the optimal oligomeric state for membrane-inserted β-barrels of protegrin-1 and related mutants. *J. Pep. Sci.* **2017**, *23*, 334−345.
(443) Sepehri, A.; Nepal, B.; Lazaridis, T. Distinct modes of action of IAPP oligomers on membranes. *J. Chem. Inf. and Mod.* **2021**, *61*, 4645−4655.
(444) Sepehri, A.; Lazaridis, T. Putative structures of membrane-embedded amyloid β oligomers. *ACS Chem. Neurosci.* **2023**, *14*, 99−110.
(445) Dutta, A.; Sepehri, A.; Lazaridis, T. Putative pore structures of amyloid β 25−35 in lipid bilayers. *Biochemistry* **2023**, *62*, 2549−2558.
(446) Maurer, M.; Lazaridis, T. Transmembrane β-barrel models of α-synuclein oligomers. *J. Chem. Inf. and Mod.* **2023**, *63*, 7171−7179.
(447) Alford, R. F.; Fleming, P. J.; Fleming, K. G.; Gray, J. J. Protein structure prediction and design in a biologically realistic implicit membrane. *Biophys. J.* **2020**, *118*, 2042−2055.
(448) Nepal, B.; Leveritt, J.; Lazaridis, T. Membrane curvature sensing by amphipathic helices: insights from implicit membrane modeling. *Biophys. J.* **2018**, *114*, 2128−2141.
(449) Zhan, H.; Lazaridis, T. Inclusion of lateral pressure/curvature stress effects in implicit membrane models. *Biophys. J.* **2013**, *104*, 643−654.
(450) Nepal, B.; Sepehri, A.; Lazaridis, T. Mechanisms of negative membrane curvature sensing and generation by ESCRT III subunit Snf7. *Protein Sci.* **2020**, *29*, 1473−1485.
(451) Nepal, B.; Sepehri, A.; Lazaridis, T. Mechanism of negative membrane curvature generation by I-BAR domains. *Structure* **2021**, *29*, 1440−1452.
(452) Vasquez Rodriguez, S. Y.; Lazaridis, T. Simulations suggest a scaffolding mechanism of membrane deformation by the caveolin 8S complex. *Biophys. J.* **2023**, *122*, 4082−4090.
(453) Tanizaki, S.; Feig, M. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* **2005**, *122*, 124706.
(454) Sayadi, M.; Tanizaki, S.; Feig, M. Effect of membrane thickness on conformational sampling of phospholamban from computer simulations. *Biophys. J.* **2010**, *98*, 805−814.
(455) Mirjalili, V.; Feig, M. Interactions of amino acid side-chain analogs within membrane environments. *J. Phys. Chem. B* **2015**, *119*, 2877−2885.
(456) Dutagaci, B.; Sayadi, M.; Feig, M. Heterogeneous dielectric generalized Born model with a van der Waals term provides improved association energetics of membrane-embedded transmembrane helices. *J. Comput. Chem.* **2017**, *38*, 1308−1320.
(457) Panahi, A.; Feig, M. Dynamic heterogeneous dielectric generalized Born (DHDGB): an implicit membrane model with a dynamically varying bilayer thickness. *J. Chem. Theory Comp.* **2013**, *9*, 1709−1719.
(458) Spassov, V. Z.; Yan, L.; Szalma, S. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B* **2002**, *106*, 8726−8738.
(459) Stern, H. A.; Feller, S. E. Calculation of the dielectric permittivity profile for a nonuniform system: application to a lipid bilayer simulation. *J. Chem. Phys.* **2003**, *118*, 3401−3412.
(460) Feig, M.; Im, W.; Brooks, C. L., III. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* **2004**, *120*, 903−911.
(461) Jaskierny, A. J.; Panahi, A.; Feig, M. Effect of flanking residues on the conformational sampling of the internal fusion peptide from Ebola virus. *PROTEINS: Struct., Func. and Bioinf.* **2011**, *79*, 1109−1117.
(462) Panahi, A.; Feig, M. Conformational sampling of influenza fusion peptide in membrane bilayers as a function of termini and protonation states. *J. Phys. Chem. B* **2010**, *114*, 1407−1416.
(463) Dutagaci, B.; Feig, M. Determination of hydrophobic lengths of membrane proteins with the hdgb implicit membrane model. *J. Chem. Inf. and Mod.* **2017**, *57*, 3032−3042.
(464) Brocke, S. A.; Degen, A.; MacKerell, A. D., Jr.; Dutagaci, B.; Feig, M. Prediction of membrane permeation of drug molecules by combining an implicit membrane model with machine learning. *J. Chem. Inf. and Mod.* **2019**, *59*, 1147−1162.
(465) Dutagaci, B.; Wittayanarakul, K.; Mori, T.; Feig, M. Discrimination of native-like states of membrane proteins with implicit membrane-based scoring functions. *J. Chem. Theory Comp.* **2017**, *13*, 3049−3059.
(466) Dutagaci, B.; Heo, L.; Feig, M. Structure refinement of membrane proteins via molecular dynamics simulations. *PROTEINS: Struct., Func. and Bioinf.* **2018**, *86*, 738−750.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(467) Tanizaki, S.; Feig, M. Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model. *The Journal of Physical Chemistry B* **2006**, *110*, 548−556.
(468) Feig, M.; Chocholoušová, J.; Tanizaki, S. Extending the horizon: towards the efficient modeling of large biomolecular complexes in atomic detail. *Theor. Chem. Acc.* **2006**, *116*, 194−205.
(469) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *PROTEINS: Struct., Func. and Bioinf.* **2010**, *78*, 1266−1281.
(470) Heo, L.; Feig, M. One bead per residue can describe all-atom protein structures. *Structure* **2024**, *32*, 97−111.
(471) Cheng, Y. M.; Gopal, S. M.; Law, S. M.; Feig, M. Molecular dynamics trajectory compression with a coarse-grained model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2012**, *9*, 476−486.
(472) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Predeus, A.; Feig, M. PRIMO: a transferable coarse-grained force field for proteins. *J. Chem. Theory Comp.* **2013**, *9*, 3769−3788.
(473) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(474) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Panahi, A.; Feig, M. Transferring the PRIMO coarse-grained force field to the membrane environment: simulations of membrane proteins and helix−helix association. *J. Chem. Theory Comp.* **2014**, *10*, 3459−3472.
(475) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(476) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comp.* **2012**, *8*, 3257−3273.
(477) Kar, P.; Feig, M. In *Advances in Protein Chemistry and Structural Biology*; Karabencheva-Christova, T., Ed.; Academic Press: 2014; Vol. *96*, pp 143−180.
(478) Feig, M.; Gopal, S. M.; Vadivel, K.; Stumpff-Kane, A. Multiscale approaches to protein modeling: structure prediction. In *Dynamics, Thermodynamics and Macromolecular Assemblies*; Kolinski, A., Ed.; Springer: New York, 2011; pp 85−109.
(479) Predeus, A. V.; Gul, S.; Gopal, S. M.; Feig, M. Conformational Sampling of Peptides in the Presence of Protein Crowders from AA/CG-Multiscale Simulations. *J. Phys. Chem. B* **2012**, *116*, 8610−8620.
(480) Kar, P.; Feig, M. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J. Chem. Theory Comp.* **2017**, *13*, 5753−5765.
(481) Chothia, C.; Levitt, M.; Richardson, D. Helix to helix packing in proteins. *J. Mol. Biol.* **1981**, *145*, 215−250.
(482) Im, W.; Lee, J.; Kim, T.; Rui, H. Novel free energy calculations to explore mechanisms and energetics of membrane protein structure and function. *J. Comput. Chem.* **2009**, *30*, 1622−1633.
(483) Lee, J.; Ham, S.; Im, W. Beta-hairpin restraint potentials for calculations of potentials of mean force as a function of beta-hairpin tilt, rotation, and distance. *J. Comput. Chem.* **2009**, *30*, 1334−1343.
(484) Lee, J.; Im, W. Implementation and application of helix-helix distance and crossing angle restraint potentials. *J. Comput. Chem.* **2007**, *28*, 669−680.
(485) Matthews, E. E.; Zoonens, M.; Engelman, D. M. Dynamic helix interactions in transmembrane signaling. *Cell* **2006**, *127*, 447−450.
(486) Schneider, D.; Engelman, D. M. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J. Mol. Biol.* **2004**, *343*, 799−804.
(487) de Planque, M. R.; Killian, J. A. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. *Mol. Memb. Biol.* **2003**, *20*, 271−284.
(488) Lee, J.; Im, W. Restraint potential and free energy decomposition formalism for helical tilting. *Chem. Phys. Lett.* **2007**, *441*, 132−135.
(489) Lee, J.; Im, W. Transmembrane helix tilting: insights from calculating the potential of mean force. *Phys. Rev. Lett.* **2008**, *100*, 018103.
(490) Lee, J.; Im, W. Role of hydrogen bonding and helix-lipid interactions in transmembrane helix association. *J. Am. Chem. Soc.* **2008**, *130*, 6456−6462.
(491) Lee, J.; Chen, J.; Brooks, C. L., III; Im, W. Application of solid-state NMR restraint potentials in membrane protein modeling. *J. Magn. Reson.* **2008**, *193*, 68−76.
(492) Tycko, R.; Stewart, P. L.; Opella, S. J. Peptide plane orientations determined by fundamental and overtone nitrogen 14 NMR. *J. Am. Chem. Soc.* **1986**, *108*, 5419−5425.
(493) Woolf, T. B.; Malkin, V. G.; Malkina, O. L.; Salahub, D. R.; Roux, B. The backbone 15N chemical shift tensor of the gramicidin channel. A molecular dynamics and density functional study. *Chem. Phys. Lett.* **1995**, *239*, 186−194.
(494) Almeida, F.; Opella, S. Fd coat protein structure in membrane environments: structural dynamics of the loop between the hydrophobic trans-membrane helix and the amphipathic in-plane helix. *J. Mol. Biol.* **1997**, *270*, 481−495.
(495) Cheng, X.; Jo, S.; Qi, Y.; Marassi, F. M.; Im, W. Solid-state NMR-restrained ensemble dynamics of a membrane protein in explicit membranes. *Biophys. J.* **2015**, *108*, 1954−1962.
(496) Im, W.; Jo, S.; Kim, T. An ensemble dynamics approach to decipher solid-state NMR observables of membrane proteins. *Biochim. et Biophys. Acta Biomembranes* **2012**, *1818*, 252−262.
(497) Jo, S.; Im, W. Transmembrane helix orientation and dynamics: insights from ensemble dynamics with solid-state NMR observables. *Biophys. J.* **2011**, *100*, 2913−2921.
(498) Kim, T.; Jo, S.; Im, W. Solid-state NMR ensemble dynamics as a mediator between experiment and simulation. *Biophys. J.* **2011**, *100*, 2922−2928.
(499) Bax, A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* **2003**, *12*, 1−16.
(500) Rathinavelan, T.; Im, W. Explicit treatment of force contribution from alignment tensor using overdetermined linear equations and its application in NMR structure determination. *J. Comput. Chem.* **2007**, *28*, 1858−1864.
(501) Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **1999**, *138*, 334−342.
(502) Rathinavelan, T.; Im, W. A novel strategy to determine protein structures using exclusively residual dipolar coupling. *J. Comput. Chem.* **2008**, *29*, 1640−1649.
(503) Wereszczynski, J.; Andricioaei, I. On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16200−16205.
(504) Rupakheti, C.; Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. Statistical mechanics of polarizable force fields based on classical Drude oscillators with dynamical propagation by the dual-thermostat extended Lagrangian. *J. Chem. Phys.* **2020**, *153*, 114108.
(505) Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell, A. D., Jr. Molecular dynamics simulations of explicitly solvated Drude polarizable systems on GPUs: implementation, validation, and benchmark. *J. Comput. Chem.* **2018**, *39*, 1682−1689.
(506) Lemkul, J. A.; Roux, B.; van der Spoel, D.; MacKerell, A. D., Jr. Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model. *J. Comput. Chem.* **2015**, *36*, 1473−1479.
(507) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2*, 87−92.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(508) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of anionic lipid membranes: development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B* **2013**, *117*, 10183−10192.
(509) Han, K.; Venable, R. M.; Bryant, A. M.; Legacy, C. J.; Shen, R.; Li, H.; Roux, B.; Gericke, A.; Pastor, R. W. Graph-theoretic analysis of monomethyl phosphate clustering in ionic solutions. *J. Phys. Chem. B* **2018**, *122*, 1484−1494.
(510) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(511) Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.* **2013**, *138*, 034508.
(512) Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D., Jr.; Roux, B. Understanding the dielectric properties of liquid amides from a polarizable force field. *J. Phys. Chem. B* **2008**, *112*, 3509−3521.
(513) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B. Simulating monovalent and divalent ions in aqueous solution using a Drude polarizable force field. *J. Chem. Theory Comp.* **2010**, *6*, 774−786.
(514) Nan, Y.; MacKerell, A. D., Jr. Balancing Group I Monatomic Ion−Polar Compound Interactions for Condensed Phase Simulation in the Polarizable Drude Force Field. *J. Chem. Theory Comp.* **2024**, *20*, 3242−3257.
(515) Lin, F. Y.; Lopes, P. E.; Harder, E. D.; Roux, B.; MacKerell, A. D., Jr. Polarizable force field for molecular ions based on the classical Drude oscillator. *J. Chem. Inf. and Mod.* **2018**, *58*, 993−1004.
(516) Villa, F.; MacKerell, A. D., Jr.; Roux, B.; Simonson, T. Classical Drude polarizable force field model for methyl phosphate and its interactions with Mg(2). *J. Phys. Chem. A* **2018**, *122*, 6147−6155.
(517) Kognole, A. A.; Aytenfisu, A. H.; MacKerell, A. D., Jr. Balanced polarizable Drude force field parameters for molecular anions: phosphates, sulfates, sulfamates, and oxides. *J. Mol. Model.* **2020**, *26*, 1−11.
(518) Rupakheti, C. R.; Roux, B.; Dehez, F.; Chipot, C. Modeling induction phenomena in amino acid cation-π interactions. *Theor. Chem. Acc.* **2018**, *137*, 174.
(519) Luo, Y.; Roux, B. Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett.* **2010**, *1*, 183−189.
(520) Lin, B.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Kirkwood-Buff analysis of aqueous n-methylacetamide and acetamide solutions modeled by the CHARMM additive and Drude polarizable force fields. *J. Chem. Phys.* **2013**, *139*, 084509.
(521) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of ion−protein interactions using the Drude polarizable force-field. *J. Phys. Chem. B* **2015**, *119*, 9401−9416.
(522) Shim, J.; Zhu, X.; Best, R. B.; MacKerell, A. D., Jr. (Ala)(4)-X-(Ala)4 as a model system for the optimization of the χ_1 and χ_2 amino acid side-chain dihedral empirical force field parameters. *J. Comput. Chem.* **2013**, *34*, 593−603.
(523) Khan, H. M.; Grauffel, C.; Broer, R.; MacKerell, A. D., Jr.; Havenith, R. W.; Reuter, N. Improving the force field description of tyrosine-choline cation-π interactions: QM investigation of pPhenol-N(Me)(4)(+) interactions. *J. Chem. Theory Comp.* **2016**, *12*, 5585−5595.
(524) Khan, H. M.; MacKerell, A. D., Jr.; Reuter, N. Cation-π interactions between methylated ammonium groups and tryptophan in the CHARMM36 additive force field. *J. Chem. Theory Comp.* **2019**, *15*, 7−12.
(525) Lin, F. Y.; MacKerell, A. D., Jr. Improved modeling of halogenated ligand-protein interactions using the Drude polarizable and CHARMM additive empirical force fields. *J. Chem. Inf. and Mod.* **2019**, *59*, 215−228.
(526) Croitoru, A.; Park, S. J.; Kumar, A.; Lee, J.; Im, W.; MacKerell, A. D., Jr.; Aleksandrov, A. Additive CHARMM36 force field for nonstandard amino acids. *J. Chem. Theory Comp.* **2021**, *17*, 3554−3570.
(527) Pane, A. J.; Yu, W.; Aytenfisu, A.; Tunyi, J.; Venable, R. M.; MacKerell, A. D., Jr.; Pastor, R. W. Development of CHARMM additive potential energy parameters for α-methyl amino acids. *J. Phys. Chem. B* **2021**, *125*, 11687−11696.
(528) Lopes, P. E. M.; Lamoureux, G.; MacKerell, A. D., Jr. Polarizable empirical force field for nitrogen-containing heteroaromatic compounds based on the classical Drude oscillator. *J. Comput. Chem.* **2009**, *30*, 1821−1838.
(529) Zhu, X.; MacKerell, A. D., Jr. Polarizable empirical force field for sulfur-containing compounds based on the classical Drude oscillator model. *J. Comput. Chem.* **2010**, *31*, 2330−2341.
(530) Baker, C. M.; MacKerell, A. D., Jr. Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers. *J. Mol. Model.* **2010**, *16*, 567−576.
(531) Baker, C. M.; Lopes, P. E.; Zhu, X.; Roux, B.; MacKerell, A. D., Jr. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. *J. Chem. Theory Comp.* **2010**, *6*, 1181−1198.
(532) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr. Polarizable force field for peptides and proteins based on the classical Drude oscillator. *J. Chem. Theory Comp.* **2013**, *9*, 5430−5449.
(533) Huang, J.; MacKerell, A. D., Jr. Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)3 peptide. *Biophys. J.* **2014**, *107*, 991−997.
(534) Lemkul, J. A.; Huang, J.; MacKerell, A. D., Jr. Induced dipole-dipole interactions influence the unfolding pathways of wild-type and mutant amyloid β-peptides. *J. Phys. Chem. B* **2015**, *119*, 15574−15582.
(535) Lin, F. Y.; Huang, J.; Pandey, P.; Rupakheti, C.; Li, J.; Roux, B. T.; MacKerell, A. D., Jr. Further optimization and validation of the classical Drude polarizable protein force field. *J. Chem. Theory Comp.* **2020**, *16*, 3221−3239.
(536) Lin, F. Y.; MacKerell, A. D., Jr. Improved modeling of cation-π and Anion-Ring interactions using the Drude polarizable empirical force field for proteins. *J. Comput. Chem.* **2020**, *41*, 439−448.
(537) Aleksandrov, A.; Lin, F. Y.; Roux, B.; MacKerell, A. D., Jr. Combining the polarizable Drude force field with a continuum electrostatic Poisson-Boltzmann implicit solvation model. *J. Comput. Chem.* **2018**, *39*, 1707−1719.
(538) Aleksandrov, A.; Roux, B.; MacKerell, A. D., Jr. pK(a) calculations with the polarizable Drude force field and Poisson-Boltzmann solvation model. *J. Chem. Theory Comp.* **2020**, *16*, 4655−4668.
(539) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D., Jr. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(540) Denning, E. J.; MacKerell, A. D., Jr. Intrinsic contribution of the 2'-hydroxyl to RNA conformational heterogeneity. *J. Am. Chem. Soc.* **2012**, *134*, 2800−2806.
(541) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; Mackerell, A. D., Jr. Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comp.* **2012**, *8*, 348−362.
(542) Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell, A. D., Jr.; Nilsson, L. Additive CHARMM force field for naturally occurring modified ribonucleotides. *J. Comput. Chem.* **2016**, *37*, 896−912.
(543) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D., Jr. Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *J. Phys. Chem. B* **2011**, *115*, 580−596.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(544) Savelyev, A.; MacKerell, A. D., Jr. All-atom polarizable force field for DNA based on the classical Drude oscillator model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.
(545) Savelyev, A.; MacKerell, A. D., Jr. Balancing the interactions of ions, water and DNA in the Drude polarizable force field. *J. Phys. Chem. B* **2014**, *118*, 6742−6757.
(546) Lemkul, J. A.; Savelyev, A.; MacKerell, A. D., Jr. Induced polarization influences the fundamental forces in DNA base flipping. *J. Phys. Chem. Lett.* **2014**, *5*, 2077−2083.
(547) Savelyev, A.; MacKerell, A. D., Jr. Differential impact of the monovalent ions Li, Na, K, and Rb on DNA conformational properties. *J. Phys. Chem. Lett.* **2015**, *6*, 212−216.
(548) Savelyev, A.; MacKerell, A. D., Jr. Competition among Li, Na, K and Rb monovalent ions for DNA in molecular dynamics simulations using the additive CHARMM36 and Drude polarizable force fields. *J. Phys. Chem. B* **2015**, *119*, 4428−4440.
(549) Lemkul, J. A.; MacKerell, A. D., Jr. Polarizable force field for DNA based on the classical Drude oscillator: I. refinement using quantum mechanical base stacking and conformational energetics. *J. Chem. Theory Comp.* **2017**, *13*, 2053−2071.
(550) Lemkul, J. A.; MacKerell, A. D., Jr. Polarizable force field for DNA based on the classical Drude oscillator: II. microsecond molecular dynamics simulations of duplex DNA. *J. Chem. Theory Comp.* **2017**, *13*, 2072−2085.
(551) Lemkul, J. A.; MacKerell, A. D., Jr. Polarizable force field for RNA based on the classical Drude oscillator. *J. Comput. Chem.* **2018**, *39*, 2624−2646.
(552) Salsbury, A. M.; Lemkul, J. A. Molecular dynamics simulations of the c-kit1 promoter g-quadruplex: importance of electronic polarization on stability and cooperative ion binding. *J. Phys. Chem. B* **2019**, *123*, 148−159.
(553) Chowdhary, J.; Harder, E.; Lopes, P. E.; Huang, L.; MacKerell, A. D., Jr.; Roux, B. A polarizable force field of dipalmitoylphosphatidylcholine based on the classical Drude model for molecular dynamics simulations of lipids. *J. Phys. Chem. B* **2013**, *117*, 9142−9160.
(554) Sengul, M. Y.; MacKerell, A. D., Jr. Accurate modeling of RNA hairpins through the explicit treatment of electronic polarizability with the classical Drude oscillator force field. *J. Comput. Biophys. Chem.* **2022**, *21*, 461−471.
(555) Klauda, J. B.; Brooks, B. R.; MacKerell, A., Jr.; Venable, R. M.; Pastor, R. W. An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer. *J. Phys. Chem. B* **2005**, *109*, 5300−5311.
(556) Levine, Z. A.; Venable, R. M.; Watson, M. C.; Lerner, M. G.; Shea, J.-E.; Pastor, R. W.; Brown, F. L. H. Determination of biomembrane bending moduli in fully atomistic simulations. *J. Am. Chem. Soc.* **2014**, *136*, 13582−13585.
(557) Sodt, A. J.; Pastor, R. W. Bending free energy from simulation: correspondence of planar and inverse hexagonal lipid phases. *Biophys. J.* **2013**, *104*, 2202−2211.
(558) Venable, R. M.; Brown, F. L. H.; Pastor, R. W. Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chem. Phys. Lipids* **2015**, *192*, 60−74.
(559) Leonard, A. N.; Wang, E.; Monje-Galvan, V.; Klauda, J. B. Developing and testing of lipid force fields with applications to modeling cellular membranes. *Chem. Rev.* **2019**, *119*, 6227−6269.
(560) West, A.; Zoni, V.; Teague, W. E., Jr.; Leonard, A. N.; Vanni, S.; Gawrisch, K.; Tristram-Nagle, S.; Sachs, J. N.; Klauda, J. B. How do ethanolamine plasmalogens contribute to order and structure of neurological membranes? *J. Phys. Chem. B* **2020**, *124*, 828−839.
(561) Yuan, Y.; Yu, Y.; Klauda, J. B. Simulations of diabetic and non-diabetic peripheral nerve myelin lipid bilayers. *J. Phys. Chem. B* **2021**, *125*, 6201−6213.
(562) Yu, Y.; Klauda, J. B. Update of the CHARMM36 united atom chain model for hydrocarbons and phospholipids. *J. Phys. Chem. B* **2020**, *124*, 6797−6812.
(563) Lucker, J.; Kio, M.; Klauda, J. B. Expanding the CHARMM36 united atom chain model for the inclusion of sphingolipids. *J. Phys. Chem. B* **2024**, *128*, 4428−4439.
(564) Yu, Y.; Kramer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D., Jr.; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated optimization of the CHARMM36 lipid force field to include explicit treatment of long-range dispersion. *J. Chem. Theory Comp.* **2021**, *17*, 1562−1580.
(565) Yu, Y. L.; Kramer, A.; Venable, R. M.; Brooks, B. R.; Klauda, J. B.; Pastor, R. W. CHARMM36 lipid force field with explicit treatment of long-range dispersion: parametrization and validation for phosphatidylethanolamine, phosphatidylglycerol, and ether lipids. *J. Chem. Theory Comp.* **2021**, *17*, 1581−1595.
(566) Venable, R. M.; Kramer, A.; Pastor, R. W. molecular dynamics simulations of membrane permeability. *Chem. Rev.* **2019**, *119*, 5954−5997.
(567) Harder, E.; MacKerell, A. D., Jr.; Roux, B. Many-body polarization effects and the membrane dipole potential. *J. Am. Chem. Soc.* **2009**, *131*, 2760−2761.
(568) Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell, A. D., Jr.; Roux, B. Drude polarizable force field for molecular dynamics simulations of saturated and unsaturated zwitterionic lipids. *J. Chem. Theory Comp.* **2017**, *13*, 4535−4552.
(569) Chen, P.; Vorobyov, I.; Roux, B.; Allen, T. W. Molecular dynamics simulations based on polarizable models show that ion permeation interconverts between different mechanisms as a function of membrane thickness. *J. Phys. Chem. B* **2021**, *125*, 1020−1035.
(570) Ngo, V.; Li, H.; MacKerell, A. D., Jr.; Allen, T. W.; Roux, B.; Noskov, S. Polarization effects in water-mediated selective cation transport across a narrow transmembrane channel. *J. Chem. Theory Comp.* **2021**, *17*, 1726−1741.
(571) Yu, Y.; Venable, R. M.; Thirman, J.; Chatterjee, P.; Kumar, A.; Pastor, R. W.; MacKerell, A. D., Jr.; Roux, B.; Klauda, J. Drude polarizable lipid force field with explicit treatment of long-range dispersion: parametrization and validation for saturated and unsaturated zwitterionic lipids. *J. Chem. Theory Comp.* **2023**, *19*, 2590−2605.
(572) Hatcher, E.; Guvench, O.; Mackerell, A. D., Jr. CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates and inositol. *J. Chem. Theory Comp.* **2009**, *5*, 1315−1327.
(573) Hatcher, E.; Guvench, O.; Mackerell, A. D., Jr. CHARMM additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides, and fructofuranose. *J. Phys. Chem. B* **2009**, *113*, 12466−12476.
(574) Raman, E. P.; Guvench, O.; MacKerell, A. D., Jr. CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *J. Phys. Chem. B* **2010**, *114*, 12981−12994.
(575) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison II, F. W.; Mackerell, A. D., Jr. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comp.* **2011**, *7*, 3162−3180.
(576) Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; Mackerell, A. D., Jr. CHARMM additive all-atom force field for phosphate and sulfate linked to carbohydrates. *J. Chem. Theory Comp.* **2012**, *8*, 759−776.
(577) He, X.; Lopes, P. E. M.; MacKerell, A., Jr. Polarizable empirical force field for acyclic poly-alcohols based on the classical Drude oscillator. *Biopolymers* **2013**, *99*, 724−738.
(578) Small, M. C.; Aytenfisu, A. H.; Lin, F. Y.; He, X.; MacKerell, A. D., Jr. Drude polarizable force field for aliphatic ketones and aldehydes, and their associated acyclic carbohydrates. *J. Comp. Aided Mol. Des.* **2017**, *31*, 349−363.
(579) Jana, M.; MacKerell, A. D., Jr. CHARMM Drude polarizable force field for aldopentofuranoses and methyl-aldopentofuranosides. *J. Phys. Chem. B* **2015**, *119*, 7846−7859.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(580) Patel, D. S.; He, X.; MacKerell, A. D., Jr. Polarizable empirical force field for hexopyranose monosaccharides based on the classical Drude oscillator. *J. Phys. Chem. B* **2015**, *119*, 637−652.
(581) Yang, M.; Aytenfisu, A. H.; MacKerell, A. D., Jr. Proper balance of solvent-solute and solute-solute interactions in the treatment of the diffusion of glucose using the Drude polarizable force field. *Carb. Res.* **2018**, *457*, 41−50.
(582) Aytenfisu, A. H.; Yang, M.; MacKerell, A. D., Jr. CHARMM drude polarizable force field for glycosidic linkages involving pyranoses and furanoses. *J. Chem. Theory Comp.* **2018**, *14*, 3132−3143.
(583) Pandey, P.; Aytenfisu, A. H.; MacKerell, A. D., Jr.; Mallajosyula, S. S. Drude polarizable force field parametrization of carboxylate and n-acetyl amine carbohydrate derivatives. *J. Chem. Theory Comp.* **2019**, *15*, 4982−5000.
(584) Kognole, A. A.; Aytenfisu, A. H.; MacKerell, A. D., Jr. Extension of the CHARMM classical Drude polarizable force field to N- and O-linked glycopeptides and glycoproteins. *J. Phys. Chem. B* **2022**, *126*, 6642−6653.
(585) Ruda, A.; Aytenfisu, A. H.; Angles d'Ortoli, T.; MacKerell, A. D., Jr.; Widmalm, G. Glycosidic α-linked mannopyranose disaccharides: an NMR spectroscopy and molecular dynamics simulation study employing additive and Drude polarizable force fields. *Phys. Chem. Chem. Phys.* **2023**, *25*, 3042−3060.
(586) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D., Jr. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33*, 2451−2468.
(587) Soteras Gutierrez, I.; Lin, F. Y.; Vanommeslaeghe, K.; Lemkul, J. A.; Armacost, K. A.; Brooks, C. L., III; MacKerell, A. D., Jr. Parametrization of halogen bonds in the CHARMM general force field: Improved treatment of ligand-protein interactions. *Bioorg. Med. Chem.* **2016**, *24*, 4812−4825.
(588) Kumar, A.; Pandey, P.; Chatterjee, P.; MacKerell, A. D., Jr. Deep neural network model to predict the electrostatic parameters in the polarizable classical Drude oscillator force field. *J. Chem. Theory Comp.* **2022**, *18*, 1711−1725.
(589) Heid, E.; Fleck, M.; Chatterjee, P.; Schröder, C.; MacKerell, A. D., Jr. Toward prediction of electrostatic parameters for force fields that explicitly treat electronic polarization. *J. Chem. Theory Comp.* **2019**, *15*, 2460−2469.
(590) Lin, F. Y.; MacKerell, A. D., Jr. Polarizable empirical force field for halogen-containing compounds based on the classical Drude oscillator. *J. Chem. Theory Comp.* **2018**, *14*, 1083−1098.
(591) Lin, F. Y.; MacKerell, A. D., Jr. Do halogen-hydrogen bond donor interactions dominate the favorable contribution of halogens to ligand-protein binding? *J. Phys. Chem. B* **2017**, *121*, 6813−6821.
(592) Chatterjee, P.; Sengul, M. Y.; Kumar, A.; MacKerell, A. D., Jr. Harnessing deep learning for optimization of Lennard-Jones parameters for the polarizable classical Drude oscillator force field. *J. Chem. Theory Comp.* **2022**, *18*, 2388−2407.
(593) Rupakheti, C. R.; MacKerell, A. D., Jr.; Roux, B. Global optimization of the Lennard-Jones parameters for the Drude polarizable force field. *J. Chem. Theory Comp.* **2021**, *17*, 7085−7095.
(594) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D., Jr.; Roux, B. Optimized Lennard-Jones parameters for drug-like small molecules. *J. Chem. Theory Comp.* **2018**, *14*, 3121−3131.
(595) Orr, A. A.; Sharif, S.; Wang, J.; MacKerell, A. D., Jr. Preserving the integrity of empirical force fields. *J. Chem. Inf. and Mod.* **2022**, *62*, 3825−3831.
(596) Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.* **1996**, *7*, 119−185.
(597) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How enzymes work: analysis by modern rate theory and computer simulations. *Science* **2004**, *303*, 186−195.
(598) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3188−3209.
(599) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182−187.
(600) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.
(601) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(602) Cui, Q. Quantum mechanical methods in biochemistry and biophysics. *J. Chem. Phys.* **2016**, *145*, 140901.
(603) Nam, K.; Shao, Y.; Major, D. T.; Wolf-Watz, M. Perspectives on computational enzyme modeling: from mechanisms to design and drug development. *ACS Omega* **2024**, *9*, 7393−7412.
(604) Bash, P. A.; Field, M. J.; Karplus, M. Free energy perturbation method for chemical reactions in the condensed phase: a dynamic approach based on a combined quantum and molecular mechanical potential. *J. Am. Chem. Soc.* **1987**, *109*, 8092−8094.
(605) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(606) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(607) Thiel, W.; Voityuk, A. A. Extension of the MNDO formalism to d orbitals: Integral approximations and preliminary numerical results. *Theor. Chim. Acta.* **1992**, *81*, 391−404.
(608) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific reaction parameterization for the AM1/d hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comp.* **2007**, *3*, 486−504.
(609) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(610) Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−103.
(611) Lopez, X.; York, D. M. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.* **2003**, *109*, 149−159.
(612) Walker, R. C.; Crowley, M. F.; Case, D. A. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.* **2008**, *29*, 1019−1031.
(613) Thiel, W. *MNDO97*, version 5.0. 1998.
(614) Vanden-Eijnden, E.; Venturoli, M. Revisiting the finite temperature string for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **2009**, *130*, 194103.
(615) Woodcock, H. L., III; Hodošček, M.; Brooks, B. R. Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms. *J. Phys. Chem. A* **2007**, *111*, 5720−5728.
(616) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.
(617) Schmidt, M.; Baldridge, K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(618) Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; et al. Recent developments in the general atomic and molecular electronic structure system. *J. Chem. Phys.* **2020**, *152*, 154102.
(619) Guest, M. F.; Bush, I. J.; Van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; Van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. The GAMESS-UK electronic structure package: algorithms, developments and applications. *Mol. Phys.* **2005**, *103*, 719−747.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(620) Amos, R. D.; Albert, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laidig, K. E.; Laming, G.; et al. *CADPAC, the cambridge analytic derivatives package*. 1995.
(621) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*, revision c.01; Gaussian, Inc.: Wallingford, CT, 2016.
(622) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2012**, *2*, 242−253.
(623) Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH 3 Cl + Cl - exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.
(624) König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J. Phys. Chem. B* **2005**, *109*, 9082−9095.
(625) Das, D.; Brooks, B. R. Modeling the catalytic mechanism of adenosine kinase with QM/MM methods. *Biophys. J.* **2000**, *78*, 333A.
(626) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. Optimization of quantum mechanical molecular mechanical partitioning schemes: gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* **2002**, *117*, 10534−10547.
(627) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.
(628) Pu, J.; Gao, J.; Truhlar, D. Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A* **2004**, *108*, 5454−5463.
(629) Pu, J.; Gao, J.; Truhlar, D. Generalized hybrid orbital (GHO) method for combining ab initio Hartree−Fock wave functions with molecular mechanics. *J. Phys. Chem. A* **2004**, *108*, 632−650.
(630) Brooks, I. C. L., III; Brünger, A.; Karplus, M. Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach. *Biopolymers* **1985**, *24*, 843−865.
(631) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(632) Rossi, I.; Truhlar, D. G. Parameterization of NDDO wavefunctions using genetic algorithms. an evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.* **1995**, *233*, 231−236.
(633) Doron, D.; Major, D. T.; Kohen, A.; Thiel, W.; Wu, X. Hybrid quantum and classical simulations of the dihydrofolate reductase catalyzed hydride transfer reaction on an accurate semi-empirical potential energy surface. *J. Chem. Theory Comp.* **2011**, *7*, 3420−3437.
(634) Zhou, Y.; Pu, J. Reaction path force matching: a new strategy of fitting specific reaction parameters for semiempirical methods in combined QM/MM simulations. *J. Chem. Theory Comp.* **2014**, *10*, 3038−3054.
(635) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, T. J. How large should the QM region be in QM/MM calculations? the case of catechol O-methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381−11394.
(636) Jindal, G.; Warshel, A. Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. *J. Phys. Chem. B* **2016**, *120*, 9913−9921.
(637) Roßbach, S.; Ochsenfeld, C. Influence of coupling and embedding schemes on QM size convergence in QM/MM approaches for the example of a proton transfer in DNA. *J. Chem. Theory Comp.* **2017**, *13*, 1102−1107.
(638) Das, S.; Nam, K.; Major, D. T. Rapid convergence of energy and free energy profiles with quantum mechanical size in quantum mechanical-molecular mechanical simulations of proton transfer in DNA. *J. Chem. Theory Comp.* **2018**, *14*, 1695−1705.
(639) Mehmood, R.; Kulik, H. J. Both configuration and QM region size matter: zinc stability in QM/MM models of DNA methyltransferase. *J. Chem. Theory Comp.* **2020**, *16*, 3121−3134.
(640) Pérez-Barcia, A.; Cárdenas, G.; Nogueira, J. J.; Mandado, M. Effect of the QM size, basis set, and polarization on QM/MM interaction energy decomposition analysis. *J. Chem. Inf. and Mod.* **2023**, *63*, 882−897.
(641) Nam, K.; Gao, J. L.; York, D. M. An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theory Comp.* **2005**, *1*, 2−13.
(642) Nam, K. Acceleration of ab Initio QM/MM calculations under periodic boundary conditions by multiscale and multiple time step approaches. *J. Chem. Theory Comp.* **2014**, *10*, 4175−4183.
(643) Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J. Chem. Phys.* **2013**, *139*, 244108.
(644) Schaefer, P.; Riccardi, D.; Cui, Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* **2005**, *123*, 014905.
(645) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; et al. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.
(646) Ojeda-May, P.; Li, Y.; Ovchinnikov, V.; Nam, K. Role of protein dynamics in allosteric control of the catalytic phosphoryl transfer of insulin receptor kinase. *J. Am. Chem. Soc.* **2015**, *137*, 12454−12457.
(647) Ojeda-May, P.; Mushtaq, A. U.; Rogne, P.; Verma, A.; Ovchinnikov, V.; Grundström, C.; Dulko-Smith, B.; Sauer, U. H.; Wolf-Watz, M.; Nam, K. Dynamic connection between enzymatic catalysis and collective protein motions. *Biochem.* **2021**, *60*, 2246−2258.
(648) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L., III. Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using Non-Boltzmann Bennett reweighting schemes. *J. Chem. Theory Comp.* **2014**, *10*, 1406−1419.
(649) Riccardi, D.; Schaefer, P.; Cui, Q. pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J. Phys. Chem. B* **2005**, *109*, 17715−17733.
(650) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.
(651) Pulay, P. Improved scf convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556−560.
(652) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. Ab initio molecular dynamics: propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **2001**, *114*, 9758−9763.
(653) Herbert, J. M.; Head-Gordon, M. Curvy-steps approach to constraint-free extended-Lagrangian ab initio molecular dynamics, using atom-centered basis functions: convergence toward Born-Oppenheimer trajectories. *J. Chem. Phys.* **2004**, *121*, 11542−11556.
(654) Niklasson, A. M. N.; Steneteg, P.; Odell, A.; Bock, N.; Challacombe, M.; Tymczak, C. J.; Holmström, E.; Zheng, G.; Weber, V. Extended Lagrangian Born-Oppenheimer molecular dynamics with dissipation. *J. Chem. Phys.* **2009**, *130*, 214109.
(655) Zheng, G.; Niklasson, A. M. N.; Karplus, M. Lagrangian formulation with dissipation of Born-Oppenheimer molecular dynamics using the density-functional tight-binding method. *J. Chem. Phys.* **2011**, *135*, 044122.
(656) Pulay, P.; Fogarasi, G. Fock matrix dynamics. *Chem. Phys. Lett.* **2004**, *386*, 272−278.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(657) Herbert, J. M.; Head-Gordon, M. Accelerated, energy-conserving Born-Oppenheimer molecular dynamics via Fock matrix extrapolation. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3269−3275.
(658) Christensen, A. S.; Kubar, T.; Cui, Q.; Elstner, M. Semi-empirical quantum mechanical methods for non-covalent interactions for chemical and biochemical applications. *Chem. Rev.* **2016**, *116*, 5301−5337.
(659) Řezáč, J.; Hobza, P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J. Chem. Theory Comp.* **2012**, *8*, 141−151.
(660) Gao, J.; Devi-Kesavan, L.S.; Garcia-Viloca, M. Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. application to the nucleophilic addition reaction in haloalkane dehalogenase. *Theor. Chem. Acc.* **2003**, *109*, 133−139.
(661) Feynman, R. P.; Hibbs, A. R. *Quantum mechanics and path integrals*; McGraw-Hill: New York, 1965.
(662) Voth, G. A. Feynman path integral formulation of quantum mechanical transition state theory. *J. Phys. Chem.* **1993**, *97*, 8365−8377.
(663) Berne, B. J.; Thirumalai, D. On the simulation of quantum-systems - path integral methods. *Annu. Rev. Phys. Chem.* **1986**, *37*, 401−424.
(664) Hwang, J. K.; Chu, Z. T.; Yadav, A.; Warshel, A. Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *J. Phys. Chem.* **1991**, *95*, 8445−8448.
(665) Hwang, J. K.; Warshel, A. A quantized classical path approach for calculations of quantum mechanical rate constants. *J. Phys. Chem.* **1993**, *97*, 10053−10058.
(666) Pollock, E. L.; Ceperley, D. M. Simulation of quantum many-body systems by path-integral methods. *Phys. Rev. B* **1984**, *30*, 2555−2568.
(667) Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **1995**, *67*, 279−355.
(668) Major, D. T.; Gao, J. Implementation of the bisection sampling method in path integral simulations. *J. Mol. Graph. Model.* **2005**, *24*, 121−127.
(669) Major, D. T.; Garcia-Viloca, M.; Gao, J. Path integral simulations of proton transfer reactions in aqueous solution using combined QM/MM potentials. *J. Chem. Theory Comp.* **2006**, *2*, 236−245.
(670) Sprik, M.; Klein, M. L.; Chandler, D. Staging: a sampling technique for the Monte Carlo evaluation of path integrals. *Phys. Rev. B* **1985**, *31*, 4234−4244.
(671) Azuri, A.; Engel, H.; Doron, D.; Major, D. T. Path-integral calculations of nuclear quantum effects in model systems, small molecules, and enzymes via gradient based forward corrector algorithms. *J. Chem. Theory Comp.* **2011**, *7*, 1273−1286.
(672) Vardi-Kilshtain, A.; Major, D. T.; Kohen, A.; Engel, H.; Doron, D. Hybrid quantum and classical simulations of the formate dehydrogenase catalyzed hydride transfer reaction on an accurate semiempirical potential energy surface. *J. Chem. Theory Comp.* **2012**, *8*, 4786−4796.
(673) Roston, D.; Kohen, A.; Doron, D.; Major, D. T. Simulations of remote mutants of dihydrofolate reductase reveal the nature of a network of residues coupled to hydride transfer. *J. Comput. Chem.* **2014**, *35*, 1411−1417.
(674) Doron, D.; Stojković, V.; Gakhar, L.; Vardi-Kilshtain, A.; Kohen, A.; Major, D. T. Free energy simulations of active-site mutants of dihydrofolate reductase. *J. Phys. Chem. B* **2015**, *119*, 906−916.
(675) Major, D. T.; Gao, J. An integrated path integral and free-energy perturbation-umbrella sampling method for computing kinetic isotope effects of chemical reactions in solution and in enzymes. *J. Chem. Theory Comp.* **2007**, *3*, 949−960.
(676) Gao, J.; Major, D. T.; Fan, Y.; Lin, Y.-l.; Ma, S.; Wong, K. Y. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Humana Press: 2008; Vol. *443*, pp 37−62.
(677) Gao, J.; Wong, K.-Y.; Major, D. T.; Cembran, A.; Song, L.; Lin, Y.-l.; Fan, Y.; Ma, S. In *Quantum Tunnelling in Enzyme-Catalysed Reactions*; Allemann, R. K., Scrutton, N. S., Eds.; The Royal Society of Chemistry: 2009; Chapter 5, pp 105−134.
(678) Rubinstein, A.; Major, D. T. Catalyzing racemizations in the absence of a cofactor: the reaction mechanism in proline racemase. *J. Am. Chem. Soc.* **2009**, *131*, 8513−8521.
(679) Vardi-Kilshtain, A.; Azuri, A.; Major, D. T. Path-integral calculations of heavy atom kinetic isotope effects in condensed phase reactions using higher-order Trotter factorizations. *J. Comput. Chem.* **2012**, *33*, 435−441.
(680) Fan, Y.; Cembran, A.; Ma, S.; Gao, J. Connecting protein conformational dynamics with catalytic function as illustrated in dihydrofolate reductase. *Biochem.* **2013**, *52*, 2036−2049.
(681) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14115−14120.
(682) Roston, D.; Demapan, D.; Cui, Q. Leaving group ability observably affects transition state structure in a single enzyme active site. *J. Am. Chem. Soc.* **2016**, *138*, 7386−7394.
(683) Roston, D.; Cui, Q. Substrate and transition state binding in alkaline phosphatase analyzed by computation of oxygen isotope effects. *J. Am. Chem. Soc.* **2016**, *138*, 11946−11957.
(684) Chin, S. A. Symplectic integrators from composite operator factorizations. *Phys. Lett. A* **1997**, *226*, 344−348.
(685) Chin, S. A.; Chen, C. R. Gradient symplectic algorithms for solving the Schrödinger equation with time-dependent potentials. *J. Chem. Phys.* **2002**, *117*, 1409−1415.
(686) Chin, S. A. Quantum statistical calculations and symplectic corrector algorithms. *Phys. Rev. E* **2004**, *69*, 046118.
(687) Morrone, J. A.; Srinivasan, V.; Sebastiani, D.; Car, R. Proton momentum distribution in water: an open path integral molecular dynamics study. *J. Chem. Phys.* **2007**, *126*, 234504.
(688) Perez, A.; Tuckerman, M. E. Improving the convergence of closed and open path integral molecular dynamics via higher order trotter factorization schemes. *J. Chem. Phys.* **2011**, *135*, 064104.
(689) Engel, H.; Doron, D.; Kohen, A.; Major, D. T. Momentum distribution as a fingerprint of quantum delocalization in enzymatic reactions: open-chain path-integral simulations of model systems and the hydride transfer in dihydrofolate reductase. *J. Chem. Theory Comp.* **2012**, *8*, 1223−1234.
(690) Nomura, Y.; Roston, D.; Montemayor, E. J.; Cui, Q.; Butcher, S. E. Structural and mechanistic basis for preferential deadenylation of U6 snRNA by Usb1. *Nucleic Acids Res.* **2018**, *46*, 11488−11501.
(691) Lai, R.; Li, G.; Cui, Q. Flexibility of Binding Site is Essential to the Ca2+ Selectivity in EF-Hand Calcium-Binding Proteins. *J. Am. Chem. Soc.* **2024**, *146*, 7628−7639.
(692) Elstner, M.; Seifert, G. Density functional tight binding. *Phil. Trans. R. Soc. Lond. A* **2014**, *372*, 20120483.
(693) Gaus, M.; Cui, Q.; Elstner, M. Density functional tight binding (DFTB): Application to organic and biological molecules. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49−61.
(694) Thiel, W. Perspectives on semiempirical molecular orbital theory. *Adv. Chem. Phys.* **1996**, *93*, 703−757.
(695) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comp.* **2013**, *9*, 338−354.
(696) Kubar, T.; Elstner, M.; Cui, Q. Hybrid QM/MM methods for studying energy transduction in biomolecular machines. *Annu. Rev. Biophys.* **2023**, *52*, 525−551.
(697) Hou, G.; Zhu, X.; Elstner, M.; Cui, Q. A modified QM/MM hamiltonian with the self-consistent-charge density-functional-tight-binding theory for highly charged QM regions. *J. Chem. Theory Comp.* **2012**, *8*, 4293−4304.
(698) Rowley, C. N.; Roux, B. The solvation structure of Na^+ and K^+ in liquid water determined from high level ab initio molecular dynamics simulations. *J. Chem. Theory Comp.* **2012**, *8*, 3526−3535.
(699) Gaus, M.; Jin, H.; Demapan, D.; Christensen, A. S.; Goyal, P.; Elstner, M.; Cui, Q. DFTB3 parametrization for copper: the

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

importance of orbital angular momentum dependence of hubbard parameters. *J. Chem. Theory Comput.* **2015**, *11*, 4205−4219.
(700) Watanabe, H.; Cui, Q. Quantitative analysis and correction of QM/MM boundary artifacts in adaptive QM/MM methods. *J. Chem. Theory Comp.* **2019**, *15*, 3917−3928.
(701) The PLUMED consortium.. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Meth.* **2019**, *16*, 670−673.
(702) Roston, D.; Demapan, D.; Cui, Q. Extensive free energy simulations identify water as the base in nucleotide addition by DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 25048−25056.
(703) Maag, D.; Mast, T.; Elstner, M.; Cui, Q.; Kubar, T. **O** to **bR** transition in bacteriorhodopsin occurs through a proton hole mechanism. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, No. e2024803118.
(704) Deng, J.; Cui, Q. Second-shell residues contribute to catalysis by predominately pre-organizing the apo state in PafA. *J. Am. Chem. Soc.* **2023**, *145*, 11333−11347.
(705) Lu, X.; Ovchinnikov, V.; Roston, D. R.; Demapan, D.; Cui, Q. Regulation and plasticity of catalysis in enzymes: insights from analysis of mechanochemical coupling in myosin. *Biochemistry* **2017**, *56*, 1482−1497.
(706) Goyal, P.; Qian, H. J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular simulation of water and hydration effects in different environments: challenges and developments for DFTB based models. *J. Phys. Chem. B* **2014**, *118*, 11007−11027.
(707) Li, G.; Cui, Q. pK_a calculations with QM/MM free energy perturbations. *J. Phys. Chem. B* **2003**, *107*, 14521−14528.
(708) Cates, M. S.; Berry, M. B.; Ho, E. L.; Li, Q.; Potter, J. D.; Phillips, G. N. Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin. *Structure* **1999**, *7*, 1269−1278.
(709) Jing, Z. F.; Liu, C. W.; Qi, R.; Ren, P. Y. Many-body effect determines the selectivity for Ca^{2+} and Mg^{2+} in proteins. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E7495−E7501.
(710) Ngo, V.; da Silva, M.; Kubillus, M.; Elstner, M.; Hui, L.; Roux, B.; Cui, Q.; Salahub, D.; Noskov, S. Quantum Effects in Cation Interactions with First and Second Coordination Shell Ligands in Metalloproteins. *J. Chem. Theory Comp.* **2015**, *11*, 4992−5001.
(711) Sherwood, P.; De Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discussions* **1997**, *106*, 79−92.
(712) De Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. Zeolite structure and reactivity by combined quantum-chemical−classical calculations. *J. Phys. Chem. B* **1999**, *103*, 6133−6141.
(713) Philipp, D. M.; Friesner, R. A. Mixedab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468−1494.
(714) Amara, P.; Field, M. J. Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2003**, *109*, 43−52.
(715) Riahi, S.; Roux, B.; Rowley, C. N. QM/MM molecular dynamics simulations of the hydration of Mg (II) and Zn (II) ions. *Can. J. Chem.* **2013**, *91*, 552−558.
(716) Chen, Y.; Kale, S.; Weare, J.; Dinner, A. R.; Roux, B. Multiple time-step dual-Hamiltonian hybrid molecular dynamics - Monte Carlo canonical propagation algorithm. *J. Chem. Theory Comp.* **2016**, *12*, 1449−1458.
(717) Snyder, R.; Kim, B.; Pan, X.; Shao, Y.; Pu, J. Facilitating ab initio QM/MM free energy simulations by Gaussian process regression with derivative observations. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25134−25143.
(718) Snyder, R.; Kim, B.; Pan, X.; Shao, Y.; Pu, J. Bridging semiempirical and ab initio QM/MM potentials by Gaussian process regression and its sparse variants for free energy simulation. *J. Chem. Phys.* **2023**, *159*, 054107.
(719) Rasmussen, C. E.; Williams, C. K. *Gaussian processes for machine learning*; MIT Press: Cambridge, MA, 2006; Vol. *2*.
(720) Matthews, A. G. D. G.; van der Wilk, M.; Nickson, T.; Fujii, K.; Boukouvalas, A.; León-Villagrá, P.; Ghahramani, Z.; Hensman, J. GPflow: A Gaussian process library using TensorFlow. *J. Mach. Learn. Res.* **2017**, *18*, 1−6.
(721) Gardner, J. R.; Pleiss, G.; Bindel, D.; Weinberger, K. Q.; Wilson, A. G. GPyTorch: blackbox matrix-matrix gaussian process inference with GPU acceleration. *Advances in Neural Information Processing Systems (NeurIPS)* **2018**, *31*, 7587−7597.
(722) Pan, X.; Snyder, R.; Wang, J.-N.; Lander, C.; Wickizer, C.; Van, R.; Chesney, A.; Xue, Y.; Mao, Y.; Mei, Y.; et al. Training machine learning potentials for reactive systems: A Colab tutorial on basic models. *J. Comput. Chem.* **2024**, *45*, 638−647.
(723) Glowacki, D. R.; Orr-Ewing, A. J.; Harvey, J. N. Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: F + CD3CN treated with a parallel multi-state EVB model. *J. Chem. Phys.* **2015**, *143*, 044120.
(724) Glowacki, D. R.; Orr-Ewing, A. J.; Harvey, J. N. Product energy deposition of CN + alkane H abstraction reactions in gas and solution phases. *J. Chem. Phys.* **2011**, *134*, 214508.
(725) Harvey, J. N.; O'Connor, M.; Glowacki, D. R. *Theory and Applications of the Empirical Valence Bond Approach: From Physical Chemistry to Chemical Biology* **2017**, 93−119.
(726) Penney, W. G. The electronic structure of some polyenes and aromatic molecules. III. bonds of fractional order by the pair method. *Proc. R. Soc. Lon. A. - Math. Phys. Sci.* **1937**, *158*, 0306−0324.
(727) Pauling, L. Atomic radii and interatomic distances in metals. *J. Am. Chem. Soc.* **1947**, *69*, 542−553.
(728) Johnston, H. S.; Parr, C. Activation energies from bond energies. I. hydrogen transfer reactions. *J. Am. Chem. Soc.* **1963**, *85*, 2544−2551.
(729) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(730) Nutt, D. R.; Meuwly, M. Studying reactive processes with classical dynamics: rebinding dynamics in MbNO. *Biophys. J.* **2006**, *90*, 1191−1201.
(731) Danielsson, J.; Meuwly, M. Atomistic simulation of adiabatic reactive processes based on multi-state potential energy surfaces. *J. Chem. Theory Comp.* **2008**, *4*, 1083−1093.
(732) Nagy, T.; Yosa Reyes, J.; Meuwly, M. Multisurface adiabatic reactive molecular dynamics. *J. Chem. Theory Comp.* **2014**, *10*, 1366−1375.
(733) Schmid, M. H.; Das, A. K.; Landis, C. R.; Meuwly, M. Multi-state VALBOND for atomistic simulations of hypervalent molecules, metal complexes, and reactions. *J. Chem. Theory Comp.* **2018**, *14*, 3565−3578.
(734) Landis, C. R.; Cleveland, T.; Firman, T. K. Valence bond concepts applied to the molecular mechanics description of molecular shapes. 3. applications to transition metal alkyls and hydrides. *J. Am. Chem. Soc.* **1998**, *120*, 2641−2649.
(735) Firman, T. K.; Landis, C. R. Valence bond concepts applied to the molecular mechanics description of molecular shapes. 4. transition metals with pi-bonds. *J. Am. Chem. Soc.* **2001**, *123*, 11728−11742.
(736) Tubert-Brohman, I.; Schmid, M.; Meuwly, M. A molecular mechanics force field for octahedral organometallic compounds with inclusion of the trans influence. *J. Chem. Theory Comp.* **2009**, *5*, 530−539.
(737) Lammers, S.; Lutz, S.; Meuwly, M. Reactive force fields for proton transfer dynamics. *J. Comput. Chem.* **2008**, *29*, 1048−1063.
(738) Mackeprang, K.; Xu, Z.-H.; Maroun, Z.; Meuwly, M.; Kjaergaard, H. G. Spectroscopy and dynamics of double proton transfer in formic acid dimer. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24654−24662.
(739) Xu, Z.-H.; Meuwly, M. Vibrational Spectroscopy and Proton Transfer Dynamics in Protonated Oxalate. *J. Phys. Chem. A* **2017**, *121*, 5389−5398.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(740) Karandashev, K.; Xu, Z.-H.; Meuwly, M.; Vanicek, J.; Richardson, J. O. Kinetic isotope effects and how to describe them. *Struct. Dyn.* **2017**, *4*, 061501.
(741) Xu, Z.-H.; Meuwly, M. Multistate reactive molecular dynamics simulations of proton diffusion in water clusters and in the bulk. *J. Phys. Chem. B* **2019**, *123*, 9846−9861.
(742) Yosa Reyes, J.; Nagy, T.; Meuwly, M. Competitive reaction pathways in vibrationally induced photodissociation of H_2SO_4. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18533−18544.
(743) Yosa Reyes, J.; Brickel, S.; Unke, O. T.; Meuwly, M. HSO_3Cl: a prototype molecule for studying OH-stretching overtone induced photodissociation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 6780−6788.
(744) Brickel, S.; Meuwly, M. OH-stretching overtone induced dynamics in HSO_3F from reactive molecular dynamics simulation. *J. Phys. Chem. A* **2017**, *121*, 5079−5087.
(745) Brickel, S.; Meuwly, M. Molecular determinants for rate acceleration in the claisen rearrangement reaction. *J. Phys. Chem. B* **2019**, *123*, 448−456.
(746) Rivero, U.; Unke, O. T.; Meuwly, M.; Willitsch, S. Reactive atomistic simulations of Diels-Alder reactions: The importance of molecular rotations. *J. Chem. Phys.* **2019**, *151*, 104301.
(747) Soloviov, M.; Das, A. K.; Meuwly, M. Structural interpretation of metastable states in myoglobin-NO. *Angew. Chem., Int. Ed.* **2016**, *55*, 10126−10130.
(748) Das, A. K.; Meuwly, M. Kinetic analysis and structural interpretation of competitive ligand binding for NO dioxygenation in truncated hemoglobin N. *Angew. Chem., Int. Ed.* **2018**, *57*, 3509−3513.
(749) Pearlman, D. A.; Kollman, P. A. The lag between the Hamiltonian and the system configuration in free energy perturbation calculations. *J. Chem. Phys.* **1989**, *91*, 7831−7839.
(750) Hunter, J. E.; Reinhardt, W. P.; Davis, T. F. A finite-time variational method for determining optimal paths and obtaining bounds on free energy changes from computer simulations. *J. Chem. Phys.* **1993**, *99*, 6856−6864.
(751) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(752) Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481−1487.
(753) Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.* **2001**, *114*, 7330−7337.
(754) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Chem. Phys.* **1976**, *22*, 245−268.
(755) Lead optimization in Orion using relative binding free energy (accessed August 5, 2024); https://web.archive.org/web/20240116121109, https://www.eyesopen.com/relative-binding-free-energy.
(756) Yang, W.; Cui, Q.; Min, D.; Li, H. QM/MM alchemical free energy simulations: challenges and recent developments. *Annu. Rep. Comput. Chem.* **2010**, *6*, 51−62.
(757) Heimdal, J.; Ryde, U. Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12592−12604.
(758) Hudson, P. S.; Woodcock, H. L., III; Boresch, S. Use of nonequilibrium work methods to compute free energy differences between molecular mechanical and quantum mechanical representations of molecular systems. *J. Phys. Chem. Lett.* **2015**, *6*, 4850−4856.
(759) Kearns, F. L.; Hudson, P. S.; Woodcock, H. L., III; Boresch, S. Computing converged free energy differences between levels of theory via nonequilibrium work methods: Challenges and opportunities. *J. Comput. Chem.* **2017**, *38*, 1376−1388.
(760) Schöller, A.; Kearns, F.; Woodcock, H. L., III; Boresch, S. Optimizing the calculation of free energy differences in non-equilibrium work SQM/MM switching simulations. *J. Phys. Chem. B* **2022**, *126*, 2798−2811.
(761) Schöller, A.; Woodcock, H. L., III; Boresch, S. Exploring routes to enhance the calculation of free energy differences via non-equilibrium work SQM/MM switching simulations using hybrid charge intermediates between MM and SQM levels of theory or non-linear switching schemes. *Molecules* **2023**, *28*, 4006.
(762) Boresch, S.; Woodcock III, H. L.; Hudson, P. S. Carrying out PERT/MSCALE nonequilibrium switching simulations with CHARMM. (accessed August 5, 2024);.
(763) Roux, B. The membrane potential and its representation by a constant electric field in computer simulations. *Biophys. J.* **2008**, *95*, 4205−4216.
(764) Khalili-Araghi, F.; Jogini, V.; Yarov-Yarovoy, V.; Tajkhorshid, E.; Roux, B.; Schulten, K. Calculation of the gating charge for the Kv1.2 voltage-activated potassium channel. *Biophys. J.* **2010**, *98*, 2189−2198.
(765) Li, Q.; Wanderling, S.; Paduch, M.; Medovoy, D.; Singharoy, A.; McGreevy, R.; Villalba-Galea, C. A.; Hulse, R. E.; Roux, B.; Schulten, K.; et al. Structural mechanism of voltage-dependent gating in an isolated voltage-sensing domain. *Nat. Struct. Mol. Biol.* **2014**, *21*, 244−252.
(766) Tan, K.; Sather, A.; Robertson, J. L.; Moy, S.; Roux, B.; Joachimiak, A. Structure and electrostatic property of cytoplasmic domain of ZntB transporter. *Protein Sci.* **2009**, *18*, 2043−2052.
(767) Luo, Y.; Egwolf, B.; Walters, D. E.; Roux, B. Ion selectivity of alpha-hemolysin with a beta-cyclodextrin adapter. I. Single ion potential of mean force and diffusion coefficient. *J. Phys. Chem. B* **2010**, *114*, 952−958.
(768) Egwolf, B.; Luo, Y.; Walters, D. E.; Roux, B. Ion selectivity of alpha-hemolysin with beta-cyclodextrin adapter. II. multi-ion effects studied with grand canonical Monte Carlo/Brownian dynamics simulations. *J. Phys. Chem. B* **2010**, *114*, 2901−2909.
(769) Robertson, J. L.; Palmer, L. G.; Roux, B. Multi-ion distributions in the cytoplasmic domain of inward rectifier potassium channels. *Biophys. J.* **2012**, *103*, 434−443.
(770) Roux, B. Exploring the ion selectivity properties of a large number of simplified binding site models. *Biophys. J.* **2010**, *98*, 2877−2885.
(771) Yu, H.; Noskov, S. Y.; Roux, B. Two mechanisms of ion selectivity in protein binding sites. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20329−20334.
(772) Sompornpisut, P.; Roux, B.; Perozo, E. Structural refinement of membrane proteins by restrained molecular dynamics and solvent accessibility data. *Biophys. J.* **2008**, *95*, 5349−5361.
(773) Vasquez, V.; Sotomayor, M.; Cortes, D. M.; Roux, B.; Schulten, K.; Perozo, E. Three-dimensional architecture of membrane-embedded MscS in the closed conformation. *J. Mol. Biol.* **2008**, *378*, 55−70.
(774) Sezer, D.; Freed, J. H.; Roux, B. Multifrequency electron spin resonance spectra of a spin-labeled protein calculated from molecular dynamics simulations. *J. Am. Chem. Soc.* **2009**, *131*, 2597−2605.
(775) Roux, B.; Islam, S. M. Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy. *J. Phys. Chem. B* **2013**, *117*, 4733−4739.
(776) Islam, S. M.; Stein, R. A.; McHaourab, H. S.; Roux, B. Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme. *J. Phys. Chem. B* **2013**, *117*, 4740−4754.
(777) Islam, S. M.; Roux, B. Simulating the distance distribution between spin-labels attached to proteins. *J. Phys. Chem. B* **2015**, *119*, 3901−3911.
(778) Raghuraman, H.; Islam, S. M.; Mukherjee, S.; Roux, B.; Perozo, E. Dynamics transitions at the outer vestibule of the KcsA potassium channel during gating. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1831−1836.
(779) Kazmier, K.; Sharma, S.; Quick, M.; Islam, S. M.; Roux, B.; Weinstein, H.; Javitch, J. A.; McHaourab, H. S. Conformational dynamics of ligand-dependent alternating access in LeuT. *Nat. Struct. Mol. Biol.* **2014**, *21*, 472−479.
(780) Kazmier, K.; Sharma, S.; Islam, S. M.; Roux, B.; Mchaourab, H. S. Conformational cycle and ion-coupling mechanism of the Na

+/hydantoin transporter Mhp1. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14752−14757.
(781) Shen, R.; Han, W.; Fiorin, G.; Islam, S. M.; Schulten, K.; Roux, B. Structural refinement of proteins by restrained molecular dynamics simulations with non-interacting molecular fragments. *PLoS Comput. Biol.* **2015**, *11*, No. e1004368.
(782) Vargas, E.; Bezanilla, F.; Roux, B. In search of a consensus model of the resting state of a voltage-sensing domain. *Neuron* **2011**, *72*, 713−720.
(783) Vargas, E.; Yarov-Yarovoy, V.; Khalili-Araghi, F.; Catterall, W. A.; Klein, M. L.; Tarek, M.; Lindahl, E.; Schulten, K.; Perozo, E.; Bezanilla, F.; et al. An emerging consensus on voltage-dependent gating from computational modeling and molecular dynamics simulations. *J. Gen. Physiol.* **2012**, *140*, 587−594.
(784) Bullerjahn, J. T.; von Bülow, S.; Heidari, M.; Hénin, J.; Hummer, G. Unwrapping NPT simulations to calculate diffusion coefficients. *J. Chem. Theory Comp.* **2023**, *19*, 3406−3417.
(785) Maginn, E. J.; Messerly, R. A.; Carlson, D. J.; Roe, D. R.; Elliot, J. R. Best practices for computing transport properties 1. self-diffusivity and viscosity from equilibrium molecular dynamics. *Living Journal of Computational Molecular Science* **2019**, *1*, 6324.
(786) Allen, M.; Tildesley, D. *Computer simulations of liquids*; Clarendon Press: Oxford, U.K., 1987.
(787) Roe, D. R.; Cheatham, T. E., III. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comp.* **2013**, *9*, 3084−3095.
(788) Venable, R. M.; Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr.; Pastor, R. W. Comparing simulated and experimental translation and rotation constants: range of validity for viscosity scaling. *J. Phys. Chem. B* **2010**, *114*, 12501−12507.
(789) von Bülow, S.; Bullerjahn, J. T.; Hummer, G. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure. *J. Chem. Phys.* **2020**, *153*, 021101.
(790) Yeh, I. C.; Hummer, G. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B* **2004**, *108*, 15873−15879.
(791) Saffman, P. G.; Delbruck, M. Brownian motion in biological membranes. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3111−3113.
(792) Camley, B. A.; Lerner, M. G.; Pastor, R. W.; Brown, F. L. H. Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. *J. Chem. Phys.* **2015**, *143*, 243113.
(793) Park, S.; Im, W.; Pastor, R. W. Developing initial conditions for simulations of asymmetric membranes: a practical recommendation. *Biophys. J.* **2021**, *120*, 5041−5059.
(794) Irving, J. H.; Kirkwood, J. G. The statistical mechanical theory of transport processes. IV. the equations of hydrodynamics. *J. Chem. Phys.* **1950**, *18*, 817−829.
(795) Schofield, P.; Henderson, J. R.; Rowlinson, J. S. Statistical mechanics of inhomogeneous fluids. *Proc. R. Soc. Lon. A. - Math. Phys. Sci.* **1982**, *379*, 231−246.
(796) Canham, P. B. The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *J. Theor. Biol.* **1970**, *26*, 61−81.
(797) Helfrich, W. Elastic properties of lipid bilayers - theory and possible experiments. *Zeitschrift Fur Naturforschung C* **1973**, *28*, 693−703.
(798) Campelo, F.; Arnarez, C.; Marrink, S. J.; Kozlov, M. M. Helfrich model of membrane bending: from Gibbs theory of liquid interfaces to membranes as thick anisotropic elastic layers. *Adv. Colloid Interface Sci.* **2014**, *208*, 25−33.
(799) Hossein, A.; Deserno, M. Spontaneous curvature, differential stress, and bending modulus of asymmetric lipid membranes. *Biophys. J.* **2020**, *118*, 624−642.
(800) Kozlov, M. M.; Winterhalter, M. Elastic-moduli for strongly curved monolayers - position of the neutral surface. *J. Phys. II* **1991**, *1*, 1077−1084.
(801) Sonne, J.; Hansen, F. Y.; Peters, G. H. Methodological problems in pressure profile calculations for lipid bilayers. *J. Chem. Phys.* **2005**, *122*, 124903.
(802) Sega, M.; Fábián, B.; Jedlovszky, P. Pressure profile calculation with mesh Ewald methods. *J. Chem. Theory Comp.* **2016**, *12*, 4509−4515.
(803) Lindahl, E.; Edholm, O. Spatial and energetic-entropic decomposition of surface tension in lipid bilayers from molecular dynamics simulations. *J. Chem. Phys.* **2000**, *113*, 3882−3893.
(804) Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D., Jr.; Pastor, R. W.; Klauda, J. B. CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature. *Biophys. J.* **2014**, *107*, 134−145.
(805) Hamaneh, M. B.; Buck, M. Acceptable protein and solvent behavior in primary hydration shell simulations of hen lysozyme. *Biophys. J.* **2007**, *92*, L49−L51.
(806) Hamaneh, M. B.; Buck, M. Refinement of the primary hydration shell model for molecular dynamics simulations of large proteins. *J. Comput. Chem.* **2009**, *30*, 2635−2644.
(807) Wagoner, J. A.; Pande, V. S. Finite domain simulations with adaptive boundaries: Accurate potentials and nonequilibrium move-sets. *J. Chem. Phys.* **2013**, *139*, 234114.
(808) Wu, X.; Brooks, B. R. Using the isotropic periodic sum method to calculate long-range interactions of heterogeneous systems. *J. Chem. Phys.* **2008**, *129*, 154115.
(809) Halle, B. Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. Lond. B* **2004**, *359*, 1207−1223.
(810) Steinbach, P. J.; Brooks, B. R. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9135−9139.
(811) Ravikumar, K. M.; Hwang, W. Role of Hydration Force in the Self-Assembly of Collagens and Amyloid Steric Zipper Filaments. *J. Am. Chem. Soc.* **2011**, *133*, 11766−11773.
(812) Teng, X.; Hwang, W. Effect of methylation on local mechanics and hydration structure of DNA. *Biophys. J.* **2018**, *114*, 1791−1803.
(813) Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **2023**, *32*, No. e4792.
(814) Shi, J.; Cho, J.-H.; Hwang, W. Heterogeneous and Allosteric Role of Surface Hydration for Protein−Ligand Binding. *J. Chem. Theory Comp.* **2023**, *19*, 1875−1887.
(815) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit [7] uril. *J. Chem. Phys.* **2012**, *137*, 044101.
(816) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37*, 2029−2037.
(817) Karplus, M.; Ichiye, T.; Pettitt, B. Configurational entropy of native proteins. *Biophys. J.* **1987**, *52*, 1083−1085.
(818) Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289−6292.
(819) King, B.; Tidor, B. MIST: maximum information spanning trees for dimension reduction of biological data sets. *Bioinformatics* **2009**, *25*, 1165−1172.
(820) King, B.; Silver, N.; Tidor, B. Efficient Calculation of Molecular Configurational Entropies Using an Information Theoretic Approximation. *J. Phys. Chem. B* **2012**, *116*, 2891−2904.
(821) Shi, J.; Shen, Q.; Cho, J.-H.; Hwang, W. Entropy Hotspots for the Binding of Intrinsically Disordered Ligands to a Receptor Domain. *Biophys. J.* **2020**, *118*, 2502−2512.
(822) Hwang, W.; Lang, M. J.; Karplus, M. Kinesin motility is driven by subdomain dynamics. *eLife* **2017**, *6*, No. e28948.
(823) Hwang, W.; Mallis, R. J.; Lang, M. J.; Reinherz, E. L. The $\alpha\beta$TCR mechanosensor exploits dynamic ectodomain allostery to optimize its ligand recognition site. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 21336−21345.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(824) Lange, O. F.; Grubmüller, H. Generalized correlation for biomolecular dynamics. *PROTEINS: Struct., Func. and Bioinf.* **2006**, *62*, 1053−1061.
(825) Roy, A.; Post, C. B. Detection of Long-Range Concerted Motions in Protein by a Distance Covariance. *J. Chem. Theory Comp.* **2012**, *8*, 3009−3014.
(826) Roy, A.; Hua, D. P.; Post, C. B. Analysis of Multidomain Protein Dynamics. *J. Chem. Theory Comp.* **2016**, *12*, 274−280.
(827) Brooks, B. R.; Karplus, M.; Andricioaei, I.; Archontis, G.; Best, R.; Boresch, S.; Brooks, C. L., III; Buck, M.; Caflisch, A.; Crowley, M. F.; et al. CHARMM Program Download (accessed August 5, 2024); https://academiccharmm.org/program.

https://doi.org/10.1021/acs.jpcb.4c04100
J. Phys. Chem. B XXXX, XXX, XXX−XXX