The Journal of Chemical Physics

Data-guided Multi-Map variables for ensemble refinement of molecular movies

Cite as: J. Chem. Phys. **153**, 214102 (2020); https://doi.org/10.1063/5.0022433
Submitted: 22 July 2020 . Accepted: 09 November 2020 . Published Online: 01 December 2020

John W. Vant, Daipayan Sarkar, Ellen Streitwieser, Giacomo Fiorin, Robert Skeel, Josh V. Vermaas, and Abhishek Singharoy

ARTICLES YOU MAY BE INTERESTED IN

Scalable molecular dynamics on CPU and GPU architectures with NAMD
The Journal of Chemical Physics **153**, 044130 (2020); https://doi.org/10.1063/5.0014475

Coarse graining molecular dynamics with graph neural networks
The Journal of Chemical Physics **153**, 194101 (2020); https://doi.org/10.1063/5.0026133

Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories
The Journal of Chemical Physics **153**, 134104 (2020); https://doi.org/10.1063/5.0018980

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Data-guided Multi-Map variables for ensemble refinement of molecular movies

Cite as: J. Chem. Phys. 153, 214102 (2020); doi: 10.1063/5.0022433
Submitted: 22 July 2020 • Accepted: 9 November 2020 •
Published Online: 1 December 2020

John W. Vant,[1] **Daipayan Sarkar,**[1,2] **Ellen Streitwieser,**[1] **Giacomo Fiorin,**[3] **Robert Skeel,**[4]
Josh V. Vermaas,[5,a)] **and Abhishek Singharoy**[1,a)]

AFFILIATIONS

[1] School of Molecular Sciences, Arizona State University, Tempe, Arizona 85281, USA
[2] Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47906, USA
[3] Theoretical Molecular Biophysics Laboratory, National Heart, Lung and Blood Institute, National Institutes of Health, 10 Center Drive, Bethesda, Maryland 20814, USA
[4] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona 85281, USA
[5] Computing and Computational Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA

[a)] **Authors to whom correspondence should be addressed:** vermaasjv@ornl.gov and Abhishek.Singharoy@asu.edu

ABSTRACT

Driving molecular dynamics simulations with data-guided collective variables offer a promising strategy to recover thermodynamic information from structure-centric experiments. Here, the three-dimensional electron density of a protein, as it would be determined by cryo-EM or x-ray crystallography, is used to achieve simultaneously free-energy costs of conformational transitions and refined atomic structures. Unlike previous density-driven molecular dynamics methodologies that determine only the best map-model fits, our work employs the recently developed *Multi-Map* methodology to monitor concerted movements within equilibrium, non-equilibrium, and enhanced sampling simulations. Construction of all-atom ensembles along the chosen values of the Multi-Map variable enables simultaneous estimation of average properties, as well as real-space refinement of the structures contributing to such averages. Using three proteins of increasing size, we demonstrate that biased simulation along the reaction coordinates derived from electron densities can capture conformational transitions between known intermediates. The simulated pathways appear reversible with minimal hysteresis and require only low-resolution density information to guide the transition. The induced transitions also produce estimates for free energy differences that can be directly compared to experimental observables and population distributions. The refined model quality is superior compared to those found in the Protein Data Bank. We find that the best quantitative agreement with experimental free-energy differences is obtained using medium resolution density information coupled to comparatively large structural transitions. Practical considerations for probing the transitions between multiple intermediate density states are also discussed.

I. INTRODUCTION

Single-particle cryo-electron microscopy (cryo-EM) has evolved into one of the most effective structure determination tools in modern-day structural biology. Following advances in electron detector technology,[1] and cold field-emission electron gun sources and energy filters,[2] cryo-EM has achieved resolutions rivaling those of x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy,[3] often providing novel structures or conformations.[4–6] However, static x-ray or cryo-EM structures alone offer limited information on the function of biomolecules. The determination of conformational trajectories remains a key stumbling block toward associating structure and function. These trajectories are expected to deliver substantial information beyond static structures, revealing, for example, the propagation of allosteric signals in complex biological molecules[7] and important clues to the conformational diversity of sites associated with diseases.[8,9]

Traditionally, molecular trajectories are derived using molecular dynamics (MD) simulations either by imposing classical assumptions via all-atom and coarse-grained force fields[10,11] or by introducing *ab initio* methodologies coupled with the classical particles via extended Lagrangian schemes.[12–14] However, it is now well established that biologically relevant conformational transitions and timescales remain inaccessible to brute force MD. This drawback of traditional MD has motivated the inception and application of a range of alchemical[15,16] and geometric methods[17–23] for enhanced sampling of molecular movements.

Experimental methods have also moved beyond calorimetric measures to capture the thermodynamic manifestations of structural ensembles and molecular trajectories. Single-molecule measurements routinely derive free energy profiles and rates as a function of simple distance or angular metrics,[24,25] though their spatial resolution is limited. Highly resolved molecular ensembles are determined from NMR and EPR experiments,[26,27] but such data are limited in size compared to MD and typically miss fast kinetic information. Addressing the need to construct free energy surfaces directly from experiments while simultaneously recovering the conformational changes, geometric machine learning methodologies are employed to hierarchically cluster millions of two-dimensional single-particle images onto a low-dimensional manifold using diffusion maps.[28] The population of points on this manifold is correlated with free energy changes between 10 and 100 molecular conformations by a Boltzmann factor.[29] Such an examination of the conformational trajectories (the so-called "molecular movies") from the cryo-EM data offers arguably the first experimentally verifiable and structurally resolved view of an entire free energy landscape, including both the intermediates and rare conformations. Therefore, going beyond the visualization of realistic stationary structures, incorporating these energy-ranked cryo-EM ensembles in MD can accelerate the potential of mean force (PMF) estimation from simulations.

Integration of cryo-EM data, and more generally, experimental data with MD, has followed from the development of two families of methods, namely, flexible fitting[30,31] and Bayesian inferencing.[32] While the former serves as a real space refinement tool available in almost all the structure determination software,[33,34] the latter has been successful in either folding small proteins (<115 residues)[32] or seeking small-scale structural changes of subdomains [<5 Å of root mean square deviation (RMSD)] and free energy changes within larger cryo-EM density segments.[35] The combination of flexible fitting and Bayesian inferencing[36] overcomes this system-size restriction on protein folding and captures extremely large-scale conformational transitions from cryo-EM data of heterogeneous complexes. Nevertheless, extracting the free energy from these integrative simulations is non-trivial. A reduced representation based on collective variables would lend itself to the computation of PMFs to be compared with experiments.

As a step toward facilitating data-guided free energy estimations, we propose here to use directly the 3D electron density fields to define plausible reaction coordinates. To this end, we employ the recently introduced *Multi-Map* method,[37] which uses volumetric maps to measure and simulate changes in shape for molecular aggregates. This is achieved by quantifying the similarity between the instantaneous molecular configuration and each of the target volumetric maps. The Multi-Map method has so far been successful at computing the thermodynamic cost of wetting/dewetting in hydrophobic cavities,[37] as well as membrane deformations that are spontaneous[37] or protein-induced.[38] It is thus tempting to use this method to simulate changes in the internal structure of biological macromolecules.

Here, we demonstrate that the use of volumetric maps representing cryo-EM densities resolved between 1 Å and 9 Å allows modeling of large-scale transformations in the protein structure. Biased sampling along Multi-Map variables constructed from this density induces the protein to alter its shape reversibly in the manner prescribed by the series of electron densities, and the corresponding PMF is derived from the simulated trajectory. Typically, high-quality atomistic structures for each of the relevant states are required to simulate a molecule's transformation and extract the associated PMF. However, formulating the free energy problem in terms of the density itself enables structural ensemble determination by biasing any starting model to a given state defined by a density map. By varying the maps' resolution between the atomic and molecular scale, we allow simultaneous real-space refinement of the data and biased sampling of the conformations. Thus, starting with only one high-quality atomistic structure defining one end state in a series of maps, low-resolution EM maps corresponding to the adjacent states can be sampled producing refined atomistic models for all the states. Due to such built-in refinement capability, free energy surfaces can be obtained even from Multi-Map variables based on low-resolution maps, with accuracy consistent with those from high-resolution structures and maps.

In what follows, the conformational dynamics of three protein molecules are investigated: apo and AP$_5$A-bound adenylate kinase (ADK), carbon monoxide dehydrogenase (CODH), and *Francisella* lipoprotein3 (FLPP3). To allow comparisons between the three, the synthetic density maps of equal resolutions were generated and used to construct a Multi-Map collective variable for each protein (Sec. II B). The proteins are then simulated with equilibrium MD (Sec. III A), non-equilibrium MD (Sec. III B), and enhanced sampling simulations (Sec. III C).

An analysis of the non-equilibrium work associated with these conformations offers a theoretical framework to determine the "resolvability" of a map.[39] Employing two maps for each protein, we demonstrate that a two-state Multi-Map variable is able to monitor open and closed protein conformations in equilibrium and during slow conformational transitions, and how the accuracy of the free energy estimates changes with the density map resolution. Finally, the effect of solvation environments on the PMF is discussed, and limitations in capturing nominal structural changes, such as single sidechain rearrangements, are brought to light.

II. METHODS

To quantitatively compare between the three proteins studied and explore the role of density map resolution, we generated synthetic maps from atomic models with multiple states deposited in the Protein Data Bank (PDB). These maps were then used to construct a Multi-Map variable[37] for enhanced sampling. The conformational dynamics of A → B and the reverse B → A transition were monitored using the Multi-Map variable itself, the Euclidean distance between states as measured by root mean square deviation (RMSD), sidechain contacts, and cross-correlation (CC) to the target densities. In addition to varying the system-sizes and environmental conditions, we compared the results obtained from maps generated at five different resolutions (1 Å, 3 Å, 5 Å, 7 Å, and 9 Å).

A. Molecular dynamics simulations

Modeling and system setup utilized the molecular visualization program VMD,[40] leveraging the *solvate* and *autoionize* plugins to generate a water box and NaCl neutralizing counterions for each protein system. The dimensions of the water box were chosen such that there was at least 20 Å of padding around the protein. All simulations described in this study share standard simulation parameters, consistent with the CHARMM36m and TIP3P force fields used to describe the protein, ions, and water.[41,42] These explicit solvent simulations were run with 2 fs time steps enabled by restraining bond lengths to hydrogen atoms via the SETTLE algorithm[43] and a 12 Å cutoff, which is switched at 10 Å. Temperature for all simulations was maintained by a Langevin thermostat set to 300 K, and pressure was maintained by a Langevin barostat set to 1 atm.[44,45] Long-range electrostatics were calculated using particle mesh Ewald with a 1 Å grid spacing.[46] All MD simulations in this study were carried out with the molecular dynamics engine NAMD 2.14b1,[47] which was the first version to support grid-based collective variables. The parameters for the explicit solvent simulations are outlined in Table S1. Simulation lengths and replica number are tabulated in Table S2.

B. Construction of the Multi-Map collective variable

The recently developed Multi-Map variable[37] is briefly summarized here. Given the Cartesian coordinates of N atoms of interest, indicated as $\mathbf{R} = \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ with $\mathbf{r} = (x, y, z)$ and ϕ_k $(k = 1 \ldots K)$, a set of volumetric maps, the general form of a Multi-Map variable ζ is

$$\zeta(\mathbf{R}) = \sum^{K} \xi_k \sum_{i=1}^{N} w_i \phi_k(\mathbf{r}_i) = \sum^{K} \xi_k \Phi_k(\mathbf{R}), \tag{1}$$

where w_i is the statistical weight assigned to the ith atom and ξ_k represents the contribution of the atomic configuration $\mathbf{R}$ to the state Φ_k along a K point-long pathway.[37] The physical nature of this pathway is thus determined, aside from the assigned statistical weights, by the choice of the maps themselves.

In the following, we assume each map $\phi_k(\mathbf{r})$ to represent the electron density of a protein, as it would be determined from a cryo-EM or crystallography experiment. For each 3D electron density, the value of the corresponding map $\phi_k(\mathbf{r})$ varies between a maximum value ϕ_{max} and a threshold value ϕ_{thr}, as done in the traditional implementation of the molecular dynamics flexible fitting (MDFF) method.[48] The use of ϕ_{thr} is dictated by the use of an experimentally measured map, and the choice of its value is simplified by the analysis of the cryo-EM density histogram with the MDFF plugin in VMD.[40,49,50] Generally, cryo-EM maps will display a large density peak corresponding to the solvent; a threshold value at or above the solvent peak should be chosen to yield a flat potential in the solvent regions (see the work of Wells *et al.* for a detailed discussion).[51]

To confine sampling to the transition between two states defined by cryo-EM maps, we start with two maps (ϕ_A and ϕ_B) and we choose the coefficients $\xi_A = -1$ and $\xi_B = +1$. With this choice, the Multi-Map collective variable (or colvar for short) in Eq. (1) is a two-state variable,

$$\zeta_{AB}(\mathbf{R}) = \Phi_B(\mathbf{R}) - \Phi_A(\mathbf{R}), \tag{2}$$

where $\Phi(\mathbf{R}) = \sum_i w_i \phi(\mathbf{r}_i)$ measures the fitness of the atomic configuration $\mathbf{R}$ against the three-dimensional map $\phi(\mathbf{r})$. The two-state colvar ζ_{AB} lies on a range between a minimum negative and maximum positive value, which corresponds to perfect fits to maps A and B, respectively. The range of ζ_{AB} is estimated *a priori* using the formulation described in Appendix A. Since this range can vary depending on the specifics of the system and the map resolution, during analysis, we frequently re-normalize this range into a reaction progress coordinate (changing between −1 and +1) from states A → B.

Figure 1 illustrates how the Multi-Map variable ζ is constructed to link the density maps of multiple protein states to conformational transitions, with the specific setup provided in Appendix B. The use of a two-state variable ζ_{AB} thus defined also draws upon the formalism of other two-state paradigms, such as two-state RMSDs and anisotropic networks,[52,53] where the two endpoints are known, and the transition between them is simultaneously monitored from both the pathway termini. In analogy, when ζ_{AB} is at its minimum, $\mathbf{R}$ is fitted to cryo-EM map A, while when ζ_{AB} is at its maximum, the structure, $\mathbf{R}$, is fitted to cryo-EM map B (Fig. 1). ζ_{AB} values near 0 represent protein configurations neither in state A nor in state B. The conformational space near 0 is vast, necessitating a thorough sampling of the associated cartesian space to determine any statistical average. However, the advantage of the two-state ζ_{AB} sampling protocol is that its gradient is steepest along the most direct path between states A and B. The number of these reactive conformations is much smaller than those needed to be monitored in a protocol using as a variable either Φ_A-only or Φ_B-only, where a productive

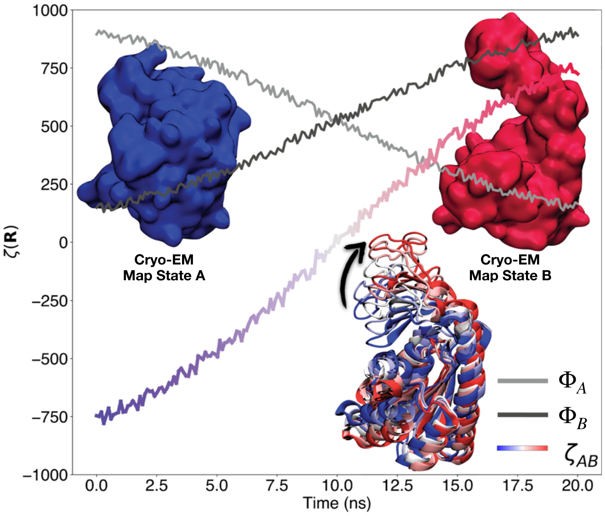

FIG. 1. Illustration of three possible choices of Multi-Map reaction coordinates capturing the closed to open transition of ADK, where Φ_A and Φ_B describes structures similarity to cryo-EM states A and B, respectively ($K = [A]$ and $\xi_A = 1$ in Eq. (1) for state A and $K = [B]$ and $\xi_B = 1$ for state B). The variable ζ_{AB} [Eq. (2)] has a negative value for structures similar to state A and a positive value for those like state B. The bottom of the plot shows snapshots of ADK transitioning from state A to B color-coded by the value of the ζ_{AB} collective variable. The arrow indicates the direction of motion the LID domain of ADK takes during the transition. Surface representations for cryo-EM maps corresponding to states A and B of ADK (PDB IDs 1AKE[97] and 4AKE,[98] respectively).

exit from A does not guarantee entry in B, and therefore, the system would need to sample an intractably larger phase space to obtain a free energy estimate.

Generalization of the two-state colvar for incorporating more than two cryo-EM maps could benefit from the existing computational methodologies to analyze density maps. If the sequence of events captured by the K maps is predetermined, e.g., by machine learning,[54] the overall conformational transition can be captured simply by concatenating the two-state transformations along the pathway.[55] If the sequence of events is unknown, then combinations of these two state transitions will have to be repeated following different orders of the events until the lowest energy or work pathway is determined for subsequent refinement.[56]

C. System preparation

Synthetic density maps were constructed for the demonstration of the Multi-Map colvar. The structures representing states A and B for each system are shown in Fig. 2 and were the density targets used to drive transitions between states. First, molecular systems were chosen based on having multiple conformational states for a single structure in the Protein Data Bank.[57] The proteins used in this study and their corresponding PDBIDs are shown in Fig. 2. These systems were translated into simulatable models through psfgen using the CHARMM36m protein force field[41] and the compatible TIP3P water model.[42] Second, a map corresponding to a specific state was generated using the `mdff sim` command, which is part of the MDFF Plugin within VMD.[40] Five maps in total were generated for each state at varying resolutions from 1 Å to 9 Å increasing by 2 Å. These density maps can be used directly without inversion to a grid potential; unlike MDFF, where density maps need to be converted to grid-potentials to be incorporated as an energy term.[48] The atoms selected to be coupled to the cryo-EM map depend on the map resolution. As a rule of thumb, based on *ab initio* electron density map refinements,[58,59] data with resolutions between 4 Å and 8 Å are fitted to backbone atoms, while resolutions higher than 4 Å are fitted to all protein atoms except hydrogen. Taking into account that we are using simulated maps, we compared the experimental $2F_c$–F_o map of ADK's closed state (PDBID 1AKE) to simulated maps at 3 Å, 5 Å, 7 Å, and 9 Å. The correlation coefficient ranged from 0.89 to 0.71, indicating that dynamics derived from the simulated and experimental maps would be similar.

Analogous to MDFF, biasing with the Multi-Map colvars requires secondary structure constraints utilizing the extraBonds feature in NAMD.[47] These constraints, which retain secondary structure folds, prevent the overfitting of the models to the maps. Additional positional and orientational constraints are used to confine the sampling in regions between adjacent cryo-EM maps, as discussed in Subsection II E.

D. Unbiased equilibrium simulation

In order to evaluate the propagation of the Multi-Map collective variable over time when not restrained, 100 ns of constant temperature and pressure sampling was performed for each of the six initial states used in the steered molecular dynamics (SMD) simulations, following the simulation parameters described above and in Table S1. These trajectories were analyzed with python scripts leveraging the python interface in VMD.[40] Due to the state drift as the simulations progressed, these analyses were carried out over the trajectory frames from 10 ns to 20 ns of simulation.

E. Steered molecular dynamics (SMD)

The formulation above lays the groundwork for using the Multi-Map colvar to define a structure's similarity to states defined by cryo-EM maps. Besides monitoring the configurational state of a protein, the Multi-Map colvar can be used to steer a protein configuration to a target map, ϕ. By employing a moving harmonic restraint, we can derive an initial pathway between states using atomic forces derived from the following equation:

$$U_{bias}(\mathbf{R}) = \frac{1}{2}\lambda(\zeta(\mathbf{R}) - \zeta_0)^2. \tag{3}$$

Here, λ is the force constant and ζ_0 is the Multi-Map colvar target value, which changes uniformly over the course of the simulation. The force constant is chosen according to the range of colvar values required for describing a transition (Table S4). The total simulation time for the SMD simulations presented here was 20 ns, and the

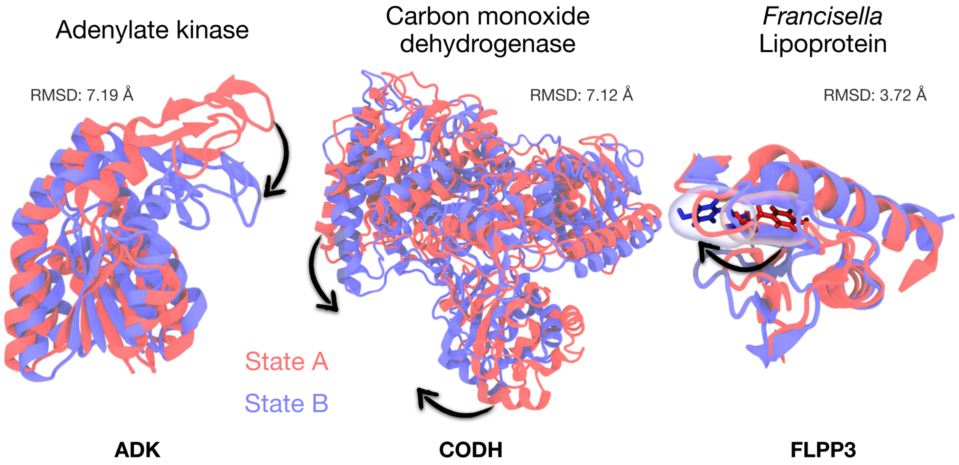

FIG. 2. The graphical overlay of states A and B for three two-state protein systems used to study the Multi-Map collective variable. Adenylate kinase (ADK) states are defined by the open (A) PDBID 4AKE[98] and closed (B), PDBID 1AKE[97] states. The carbon monoxide dehydrogenase (CODH) A and B states are taken from chains D and C of the PDBID 1OAO[90] structure. The FLPP3 A and B states are drawn from the crystal structure PDBID 6PNY[74] and the NMR structure PDBID 2MU4,[88] respectively. The flipped tyrosine residue, Tyr83, is highlighted within the FLPP3 structure.

pulling rates, as measured by center of mass movement, ranged from 0.5 Åns^{-1} to 1.5 Åns^{-1}. Expanding the harmonic and calculating the atomistic forces derived from U_{bias}, one sees that the force coming from the first term is proportional to $d\phi/dr_i$. Similar to MDFF,[48] this term localizes atoms onto the density surface. The second term, proportional to $d(\phi(\mathbf{R})\phi_0(\mathbf{R}_0))/dr_i$, is akin to taking the derivative of a correlation and acts to drive $\mathbf{R} \rightarrow \mathbf{R}_0$, effectively steering the atomic structures toward the target density ϕ_0.

The quantity and quality of cryo-EM data are typically insufficient to refine atomic models with a high degree of accuracy using ζ_{AB} as the only input. We supplement the energy function U_{bias} with additional terms used in conventional molecular dynamics, U_{ff}, the CHARMM all-atom additive potential, as well as additional constraining terms to confine sampling to realistic configurations. The following equation thus governs the simulation dynamics:

$$U_{total}(\mathbf{R}) = U_{ff}(\mathbf{R}) + U_{bias}(\mathbf{R}) + U_{pos\text{ - }ori}(\mathbf{R}) + U_{SS}(\mathbf{R}). \qquad (4)$$

Here, U_{SS} represent secondary structure restraints defined using the extraBonds feature of NAMD[47] and $U_{pos-ori}$ include center-of-mass and orientational restraints provided by the Colvars module[60] to prevent protein diffusion. U_{SS} maintain the set of secondary structure folds the systems start with and are typical for MDFF structure refinement.[49] The U_{SS} constraint could be omitted when folding from a random coil based on cryo-EM map data;[61] however, the protein folding problem remains challenging in MD for proteins larger than 115 residues.[32] Thus, we start with a model that has commensurate folds to the target states. The term $U_{pos-ori}$ ensures that the system does not translate or rotate relative to the cryo-EM maps when biasing ζ_{AB} near 0. Such treatment reduces the orthogonal degrees of freedom that do not contribute meaningfully to the transition pathway and is commonly seen in free energy perturbation simulations.[62] The positional and orientational constraints ensure that structures derived along the pathway are relevant to the states defined by the cryo-EM maps.

F. Bias exchange umbrella sampling

Enhanced sampling methods are used to calculate the free energy change between two states in molecular simulations. Some of the well-established methods are umbrella sampling (US),[17] adaptive biasing force (ABF) method,[63] and metadynamics.[64] We use an exchanging US algorithm to reconstruct the PMF along the reaction coordinate path sampled using SMD. Here, we briefly describe the method and application to the recently introduced system-specific reaction coordinate, cryo-EM map density collective variable, $\zeta(\mathbf{R})$. For biomolecular systems with large degrees of freedom, the sampling efficiency of US is significantly improved when combined with a replica-exchange scheme, hence the term bias-exchange (or replica-exchange) umbrella sampling.[65,66] In replica-exchange MD or bias exchange umbrella sampling (BEUS), each replica (or window) is assigned a different value of a given property for the system. Periodic attempts are made to exchange between replicas using a rule defined by the Metropolis criteria. The exchange rule is set based on biasing potentials, attempting a swap every 500 steps (or 0.5 ps) over a range of 126 windows. The mixing of replicas in BEUS, ensures continuous sampling for protein conformations between each replica, generating a more reliable free energy profile for the process. To remove any unphysical bias toward a particular state, 50% of the windows, i.e., 63 of them were initialized with models picked from the A → B SMD, evenly interspersed with initial models for the other 63 windows chosen from steering along the B → A direction. Further details on BEUS and applications to different biomolecular systems are discussed elsewhere.[56,65,67]

G. Calculating potentials of mean force

There are various methods of assessing the potential of mean force. For the steered molecular dynamics trajectories, the non-equilibrium work is computed internally by the Colvars module.[60] The non-equilibrium work permits an initial free energy estimate based on the second law of thermodynamics, which has as a consequence that the work for a non-equilibrium process W is bounded from below by the overall free energy difference, $W \geq \Delta F$,[68] although the short simulations typically substantially overestimate the difference. From the bias exchange umbrella sampling simulations, we estimate the free energy profiles and their uncertainties along the defined reaction coordinate using a modified version of BayesWHAM.[69] The implementation has been accelerated by using Habeck's Gibbs sampling method[70] rather than Metropolis–Hastings sampling as originally implemented.[69] As an additional check, multistate Bennett's acceptance ratio calculations,[71] as implemented in pyMBAR, are used to verify our methodology. Uncertainty estimates are obtained by trajectory subsampling to compute the variation in the computed free energies, which is used to assess convergence.

III. RESULTS AND DISCUSSION

The changes in the Multi-Map colvar were monitored during equilibrium, non-equilibrium, and free energy simulations. We focus on how these colvars track global and local conformational rearrangements. The results bring to light the pros and cons of applying this reduced representation over traditional geometric collective variables emerging from a linear combination of atomic coordinates.

A. Multi-Map colvars describe large-scale structural transitions at >5 Å density resolution

First, we seek to determine whether the colvar can distinguish between states A and B based exclusively on the ζ_{AB} value computed from Eq. (2). Figure 3(a) shows histograms of the ζ_{AB} colvar distribution during an explicit solvent equilibrium simulation. At resolutions of 3 Å–9 Å, we find that the conformations initiated at state A retain the negative values they start with, and similarly, configurations that start in the state B sample around the positive values for the ζ_{AB} coordinate. The initial values of the ζ_{AB} colvar for simulations starting in either state A or B reflect a perfect fit to the data by construction and are near unity in the scaled ζ_{AB} space. These are the minimum and maximum values the ζ_{AB} colvar can assume. For the colvars derived from high-resolution density maps, the equilibrium conformations rapidly drift away from these limiting ζ_{AB} values; the explicit values of ζ_{AB} are presented in Fig. S2. The ζ_{AB} relaxes to distributions with a near-zero mean, implying minimal separation between the states and is shown clearly through a map correlation coefficient becoming equivalent (Figs. S5–S7).

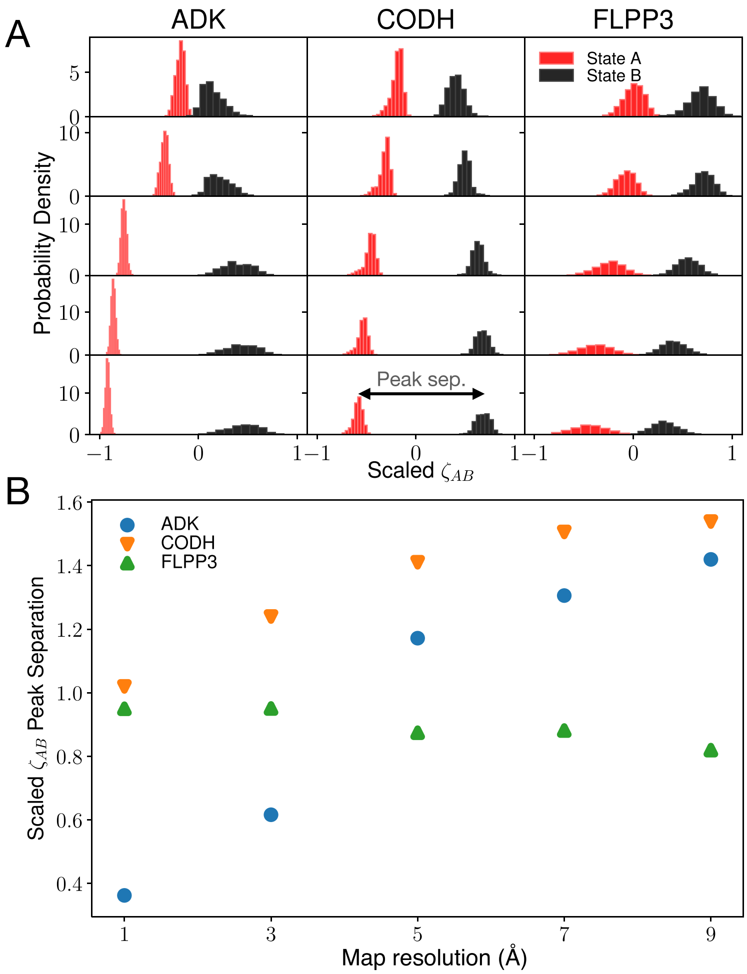

FIG. 3. (a) Histograms of scaled ζ_{AB} values from explicit solvent equilibrium simulations (Table S1 for simulation parameters) of ADK, CODH, and FLPP3 starting in either state A or B (red and black, respectively). The 10 ns simulation following a 10 ns initialization period was used to construct the histograms (see Fig. S2 for the full equilibrium trajectories). The ζ_{AB} values were normalized between −1 and 1 by dividing each value by half of the range seen during the equilibrium simulation. The double-sided arrow seen in the CODH 9 Å map plot is a visual interpretation of the peak separation. (b) Histogram maximum probability peak separation as a function of the map resolution used to derive the ζ_{AB} colvar. The peak separation values are based on the scaled ζ_{AB} profile, where the maximum separation that can be obtained is 2. For simulation lengths, see Table S2.

This trend of ζ_{AB} approaching 0 observed with all high-resolution maps also suggests the need for sampling an intractably high number of conformations for estimating averages, irrespective of the system-size. In contrast, the degeneracy in conformations underlying the endpoint ζ_{AB} values is higher for the colvars derived at a lower resolution. This is reflected in the broader distribution of low-resolution ζ_{AB} values, peaked about non-zero means (Fig. 3).

At any given resolution, the range of ζ_{AB} values visited is the highest for CODH, followed by ADK and then FLPP3. These values follow a trend guided by the number of atoms in these systems, where the range increases with increasing system size (Table S4). The separation between $\zeta_{AB}(\mathbf{R}_A)$ and $\zeta_{AB}(\mathbf{R}_B)$ is, therefore, most prominent in CODH and least in FLPP3. Conversely, for any system size, the scaled ζ_{AB} peak separation increases and finally plateaus with lower map resolution [Fig. 3(b)]. Fuzzier density features for the maps of lower resolution have reduced values of ϕ for any r_i, resulting in lower values of the ζ_{AB} summation in Eq. (2). Despite this lower range of ζ_{AB} values, the separation of states improves dramatically at lower resolutions (Fig. 3).

For resolutions of 5 Å or lower, the colvar tracks distinct large-scale conformational changes, clearly representing the open and close states in ADK and CODH [Fig. 3(a)]. At these resolutions, the separation between states on the ζ_{AB} profile is roughly equivalent to or higher than the corresponding RMSD_{AB} scaled peak separation [Fig. 3(b) and Fig. S4 and Sec. S1]. This finding suggests that the path length in the ζ-space is longer than or equivalent to the path created in the space of geometric collective variables (Fig. S3). The longer path length in the ζ can accommodate more bins to separate the meta-stable states between endpoints. A more important benefit of ζ over the traditional geometric collective variables is that knowledge of the endpoint structures is not required. Unlike geometric collective variables (such as RMSD_{AB}) where atomic models or structures need to be fit to each of the K maps to define a pathway, the definition of ζ requires only the knowledge of one endpoint (Appendix A). This formulation precludes the need for *a priori* real space refinements, although the resolution needs to be accounted for (Sec. S3).

Molecular rearrangements for both ADK and CODH require large domain movements and have significantly larger RMSDs when compared to FLPP3 (Fig. 2). In contrast, FLPP3 conformational transition involves breaking an interior hydrogen bond made by Tyr83 and the tyrosine residue's movement to an outward-facing conformation. The subtle rearrangements involved in the Tyr83 flip are only distinguishable with high-resolution maps (i.e., 1 Å and 3 Å). The lower resolution maps, and therefore the ζ_{AB} colvar, cannot distinguish between FLPP3 configurations in state A or B. At low-resolution, the cryo-EM maps for FLPP3 states A and B are highly similar with a correlation coefficient of 0.85, 0.92, and 0.95 for map resolutions 5 Å, 7 Å, and 9 Å, respectively (Figs. S2–S7).

Altogether, the ζ_{AB} colvar has two requirements to be able to discern protein configurations into individual states. First, the transition between states needs to be large enough to distinguish between cryo-EM maps at nominal resolutions. Second, the cryo-EM map needs to be at a low enough resolution to incorporate an ensemble of structures undergoing thermal motion into a single state defined by the map. This first requirement is system dependent. The second requirement can be met by low-pass filtering of high-resolution cryo-EM maps using VMD's voltools plugin if required.[31] The blurring adds Gaussian halfwidths σ to the maps and enables the maps to account for more structures in their state definitions and thus a colvar, which is better able to distinguish between the structural ensembles from states A and B.

B. Steering MD along low-resolution Multi-Map variables produces complete transitions

Trial pathways for probing large-scale conformational transitions are often generated using external forces via steered MD (SMD) simulations.[72,73] Starting from state A or B for each of the three proteins, SMD was used to drive the transition to the other endpoint using the Multi-Map colvar defined at density map resolutions between 1 Å and 9 Å. The transitions were monitored as a function of transition progress using ζ_{AB} (Fig. S8), as well as by changes in RMSD (Figs. S9 and S10), and density map correlation coefficients (Figs. S12 and S11) relative to both the initial and final states. Regardless of the preferred metric, the steered trajectories consistently demonstrate that our enforced Multi-Map biases allow the initial state to approach the final state. At resolutions 5 Å or lower, the colvar traces for the forward and reverse paths derived from two independent simulations, considerably overlap. By comparing the pathways via RMSD calculations of the A → B and B → A trajectories for three independent SMD protocols with ζ_{AB}, RMSD_{AB}, and LID-CORE distance harmonic potentials, we find that ζ_{AB} steering has marginally lower hysteresis than RMSD_{AB} steering, particularly at the endpoints, which is significantly lower than center of mass distance steering (Fig. 4). Taken together, SMD with the ζ_{AB} variable is found to achieve reproducible non-equilibrium

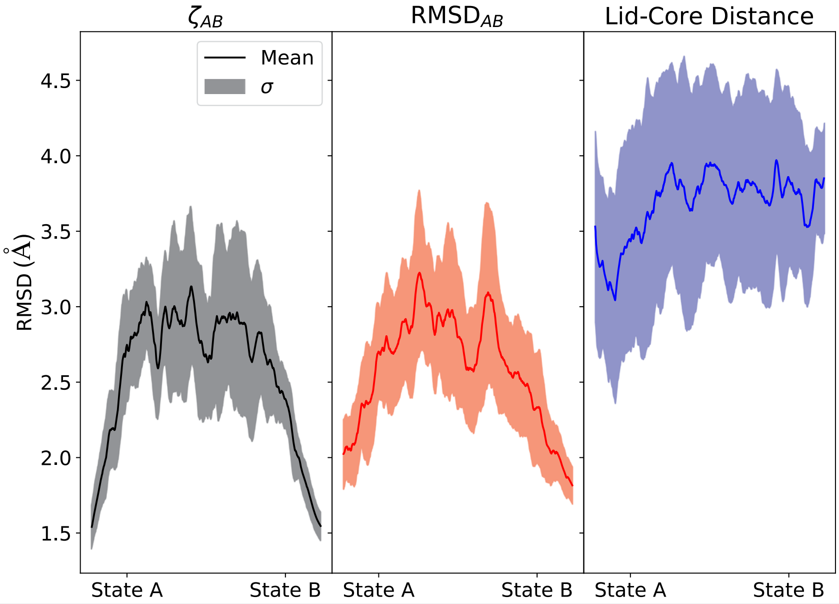

FIG. 4. Hysteresis observed in pathways described by 20 ns explicit solvent SMD simulations of the Multi-Map (left), RMSD_{AB} (center), and LID-CORE distance (right) variables. The distance between the center of mass of the LID (residues 1–29, 68–117, and 161–214) and CORE (residues 118–167) domains in ADK is defined as the LID-CORE distance. Hysteresis at any point along the conformational variable is indicated by deviations in the structures sampled in the forward and backward SMD trajectories visiting that point, which we measure using RMSD. SMD trajectories along the A → B trajectory were aligned to those from the B → A such trajectory by reversing the order of their snapshots, and overlaying the first frame of the A → B trajectory to the last frame of the trajectory from B → A trajectory. Five A to B and five B to A trajectories were used in total giving 25 RMSD calculations per frame. The shaded region represents 1 standard deviation from the mean RMSD value for that frame. Plotted is the moving average over 4000 frames with a window size of 100 frames. While the short timescale simulations do not capture microscopic reversibility, hysteresis is the least at the endpoints and magnifies in the middle of the trajectories. ζ_{AB} offers the lowest irreversibility within the simulated forward and backward trajectories. For simulation lengths and system size information, see Table S2.

pathways even within the short simulation times used (20 ns each direction).

At high-resolution (1 Å and 3 Å), ADK and FLPP3 are driven through transitions that appear to be complete along the correlation coefficient dimension but struggle to reach the extreme ζ_{AB} values that are expected for state A or B (Fig. S12). Incomplete transitions are in large part due to the well-defined local density features. At high-resolutions, 1 Å and 3 Å, the density features are quite narrow and include sidechain conformations. Thermal motion prevents simulated systems from perfectly fitting to the maps, similar to the fast correlation decay observed in equilibrium simulation (Figs. S5 and S6). Thus, for the high-resolution maps, the collective variable cannot find configurations that perfectly fit the maps and complete the transition in ζ_{AB} space, even though other collective variable metrics such as RMSD (Fig. S9) or density correlation coefficient (Fig. S11) indicate that the transition has completed.

Low-resolution maps are also not without their issues within a steered simulation context. As visualized in Fig. 2 and noted in discussions of prior simulations,[74] the transition in FLPP3 depends on the rotation of Tyr83 from packing in the interior to becoming solvent-exposed. This relatively subtle shift is difficult to capture in the context of low-resolution electron densities, unlike the much larger conformational changes for ADK and CODH (Fig. 2). For low-resolution maps, similar states have largely overlapping electron densities, resulting in comparatively few density differences that the Multi-Map collective variable can exploit to drive a conformational change. This is particularly clear in the variation of abscissa ranges in Fig. S12, where ζ_{AB} varies less for FLPP3 along the transition than it does for the other systems tested. In summary, the small structural change for FLPP3 complicates transitions driven by low-resolution structural data, while the narrow densities for high-resolution data complicate the search process for poses that do not already fit the imposed density well. Nonetheless, with density data at 5 Å or lower resolution, large-scale conformational transitions on the order of 7 Å change in RMSD are captured with minimal hysteresis between the forward and reverse pathways.

The "resolvability"[39] of a map can be inferred from the magnitude of non-equilibrium work needed to fit that map. While the lower resolution maps require less work to fit a broad set of correct models into the map, higher resolution maps require more work to fit the correct model (Fig. S14), as the number of such models nonlinearly decreases with an increase in resolution.[75] The work needed for fitting higher resolution maps further increases with system-size from FLPP3 to ADK to CODH. Establishing a common theoretical underpinning for why real-space refinement becomes more cumbersome for high-resolution maps.[76] The non-equilibrium work of the Multi-Map colvar explains several refinement challenges faced by MDFF, ROSETTA-EM, or other density-guided MD protocols when the low-resolution EM refinement tools were originally re-purposed to resolve high-resolution density maps.[31,77] A large amount of work is needed to bring structures into density features when maps A and B are non-overlapping. This scenario is prominent with sub-3 Å density maps wherein negative work accumulates as the structure initially relaxes and deviates from the small set of correct models defined by the high-resolution density map of the initial state. As the system approaches the final state, a large amount of positive work is needed to make up for the significant loss of entropy at the end state. Such physical limitations have proved detrimental in extending straight-forward MD refinements of maps between 3 Å and 5 Å resolution.[78] At lower resolution, when the overlap between maps A and B improves, less work is needed, and most of the MD is productive in fitting the model to the map. Furthermore, once the transition has completed we find that the structural statistics, as measured by MolProbity,[79] improve over the scores calculated from the deposited

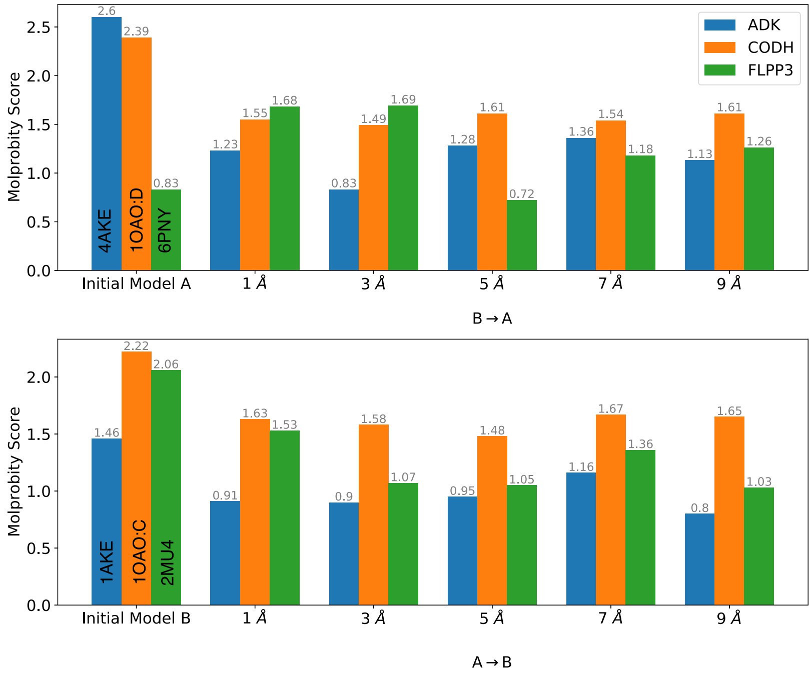

FIG. 5. MolProbity scores for the initial models of ADK, CMDH, and FLPP3, as well as those after the backward or B → A (top) and forward or A → B (bottom) SMD transitions at 1 Å, 3 Å, 5 Å, 7 Å, and 9 Å. The PDB IDs for each of the initial models are labeled on the bars. The number at the top of each bar is the calculated overall MolProbity score for that structure. Noting that a lower Molprobity score indicates higher structural quality, we find that with the exception of the 6PNY model for FLPP3, the Multi-Map colvar simulations have determined consistently higher quality structures than the submitted PDB models.

PDB structures except for the case of the FLPP3 structure, 6PNY (Fig. 5), which had incomplete transitions in the steered molecular dynamics simulations (Fig. S12). The refinement of structures is seen in their growth of correlation coefficients during the SMD (Figs. S12 and S13) and improved overall MolProbity scores for the associated structures, even when re-evaluated against deposited data from the PDB (Fig. 5).

C. Free energy profiles with Multi-Map variables resolve large-scale conformational transitions

We test how a reaction coordinate defined by the difference between the two maps estimates the relative free energy differences between the two end states. To this end, we implement the BEUS protocols to derive free energy differences along the ζ_{AB} profile for the three examples. The windows are linearly distributed along ζ_{AB} pathways of minimal hysteresis derived from Fig. S12.

1. ADK open-to-close transition

To examine the relationship between free energy differences and map resolution, BEUS was performed with ζ_{AB} at five different map resolutions for ADK [see Figs. 6(a) and 6(b)]. At the highest resolutions (1 Å and 3 Å), the endpoints are thermodynamically inaccessible, consistent with the fact that only a handful of conformations can fit the distinct features of the high-resolution density maps. Thus, there is a frustration that entropically boosts free energy at extreme ζ_{AB} values. These artificially sharp features on the energy landscape subside at the lower resolutions (≥5 Å), wherein the states close to the endpoints A and B (indicated by −1 and +1 values along the conformational coordinate), become more thermodynamically accessible. Furthermore, there is disagreement in free energy profiles between the high and low-resolution maps. While these differences are drastic for the 1 Å and 3 Å maps compared to the lower resolution maps, the 5 Å–9 Å maps are in agreement both in terms of endpoint free energy values and their differences between the local minima.

Free energy profiles of both ADK systems have been obtained in previous computational studies using geometric collective variables, such as RMSD, and the angle between domains and distance between the center of mass of one domain to that of another domain.[80–82] Free energy estimates have also been derived from single-molecule Förster resonance energy transfer (FRET)

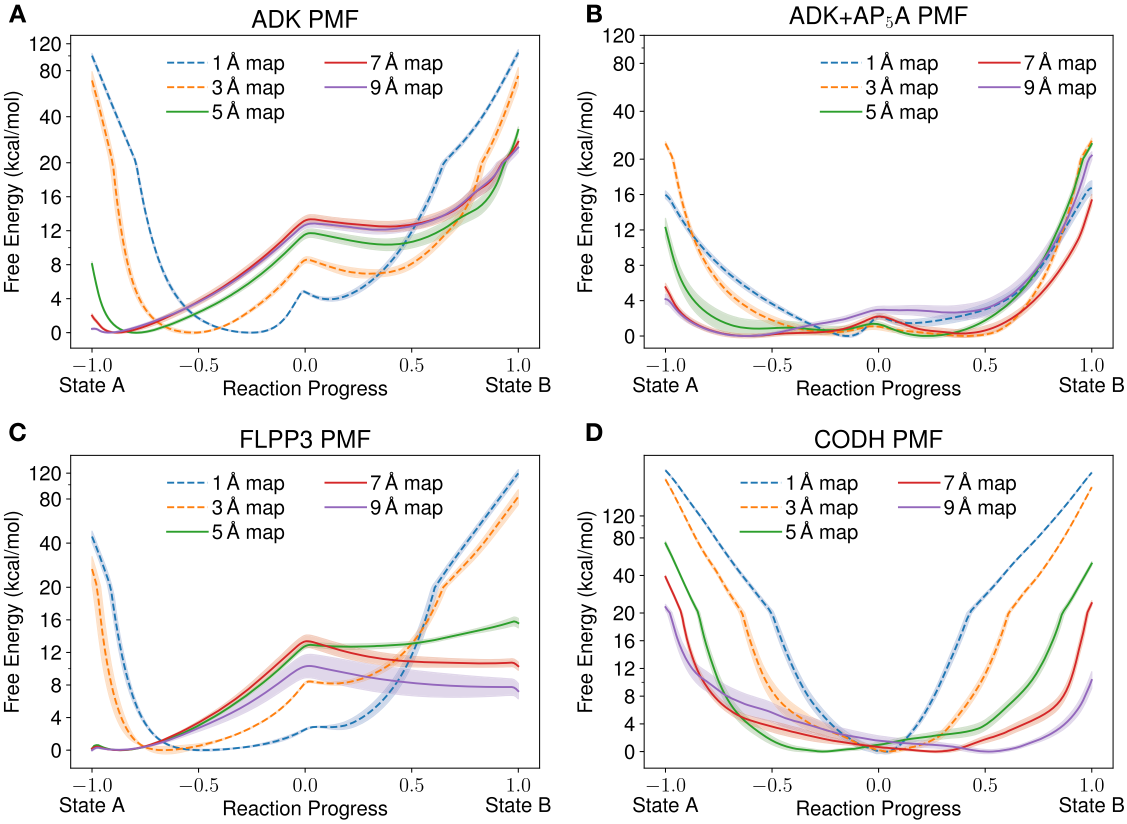

FIG. 6. Free energy profiles for ADK without (a) and with (b) the AP_5A ligand, FLPP3 (c) and CODH (d), obtained from explicit solvent bias exchange umbrella sampling simulations using the Multi-Map reaction coordinate across resolutions 1 Å, 3 Å, 5 Å, 7 Å, and 9 Å. The abscissa has been scaled so that the reaction coordinate ranges are commensurate across different resolutions. Error estimates obtained by the spread of 5 ns trajectory subsamples are represented by the shaded regions around each free energy estimate. For convergence estimates based on subsamples, see Figs. S16–S30. For simulation lengths and replica information, see Table S2.

spectroscopy and NMR studies.[83,84] Despite significant discrepancies in the magnitude of free energy estimates present in the literature, a clear trend is observed across all studies. The closed state of apo-ADK (denoted as state B here) is less favorable than the open state (A); this trend reverses for holo-ADK.[80–83] The free energy profiles for both apo and holo (AP_5A inhibited) ADK presented in Figs. 6(a) and 6(b) reproduce the expected trends, but now through the refinement of only low-resolution density information as a guide for defining a suitable collective variable, rather than geometric collective variables defined using expensive high-resolution molecular structures.

Using the density maps of resolution 5 Å or lower, apo-ADK's free energy difference between states is ~10 kcal/mol, which is comparable to the values found in previous computational studies.[80,81,85] In contrast, the holo-ADK shifts the population toward the closed conformation (state B), with both the open and closed states having approximately equivalent free energy. For map resolutions lower than 5 Å, the AP_5A binding changes the well depths for both states to be within 2 kcal/mol of one another. A result is in close agreement with experimentally observed small-molecule FRET and NMR population distributions (see Table S3).[83,84] The 5 Å and 7 Å maps for holo-ADK converge at free energy differences between basins of ~−1 kcal/mol and 0 kcal/mol, respectively, which is in the vicinity of experimental estimates of −1 kcal/mol and −0.5 kcal/mol.

Our results enhance other computational studies in terms of the magnitude of the free energy differences found (see Table S3) and the abundance of minima observed. Previous computations did not identify a local minimum for the closed state of apo-ADK.[81,85] The most analogous of these studies used RMSD_{AB} as the biasing variable[80] and was able to obtain single minima for the open state but had unrealistically high free energy values as this variable constrained the protein conformation to a small number of structures with very low RMSD values to the closed state. In addition, we found the RMSD_{AB} variable to have considerable hysteresis at the endpoints (Fig. 4). In contrast, as the Multi-Map variable approaches values corresponding to the closed state, low-resolution maps provide a constraint that encompasses a greater number of similar structures isolating a plausible thermodynamic state consistent with the experimentally observed maps. This distribution of free energy implies that induced fit (and not conformational selection) is at work to enable AP_5A binding to ADK, a result that matches kinetic assays and NMR results.[86,87] Put together, the free energy profiles obtained though the two state Multi-Map method are able to distinguish between apo and holo ADK and produce biophysically relevant information about the systems.

2. Sidechain flip in FLPP3

To elucidate the limitation of the Multi-Map colvar, we present the free energy profile for Tyr83 flip and the associated pocket opening in FLPP3 [Fig. 6(c)]. We have recently isolated the open, close (states A and B, respectively) and occluded conformations of FLPP3 using serial femtosecond crystallography (SFX) and NMR spectroscopy.[74,88] Umbrella sampling simulations enabled the hierarchization of these conformations in terms of distinct Tyr83 orientations,[74] whereby the open conformation with solvent-exposed Tyr83 was found to be significantly more stable. Already depicted in Fig. S12, the Multi-Map colvar fails to capture the complete transition between the Tyr83-flipped end states of FLPP3. Unlike ADK, where the most separated end states were observed 5 Å or lower resolutions, the most resolved FLPP3 states are seen at 3.0 Å (Fig. 3). Under these conditions, the PMF shows a more probable closed state or state A. At lower resolutions (≥5 Å), the converged statistics heavily favor state A over state B. This result is in stark contrast to NMR that favors state B to state A by a population ratio of 2:1.[74] The crystal structure (state A) seen in the SFX data is over stabilized by lattice contacts and is, therefore, even rarer. Thus, for this system, the BEUS of the Multi-Map colvar converged to unreliable results. This failure can be rationalized using Fig. S12. Unlike the ADK and CODH, the end states for FLPP3 are only accessible to the Multi-Map colvar at higher resolutions, *albeit* with hysteresis artifacts in forward and backward pathways (Fig. S8). Since the sidechain motions contribute minimally to the overall map transformation, the slow degrees of freedom of the subtle transition of Tyr83 from an outward facing conformation to an inward facing conformation were not captured by the Multi-Map colvar. We expect a slower SMD to allow more chances for the Tyr83 to flip outward and therefore address this error, possibly only at higher resolutions given the small-scale structural transition at hand.

3. Hinge-bending in CODH

The largest system we applied the Multi-Map reaction coordinate to was CODH. Unlike PMFs for ADK and FLPP3, which were relatively consistent in shape between different input map resolutions, CODH demonstrates significant changes in the shape of the free energy profile in a resolution-dependent manner [Fig. 6(d)]. For similar entropic considerations mentioned previously, the PMF minima are near $\zeta_{AB} = 0$ for 1 Å and 3 Å maps. The profiles for these high-resolution maps have sharp edges due to the difficulty of fitting thermalized states into a high-resolution map. The valley broadens when 5 Å resolution maps are applied to the Multi-Map colvar, offering comparable probability to the open and closed CODH structures, marginally skewed toward state A, while 7 Å and 9 Å maps are skewed toward state B.

This energy landscape is consistent with the biochemical knowledge available for CODH, which implies that the protein structure fluctuates to accommodate substrate ingress.[89] Indeed, the existence of both states within a single crystal structure implies that at least in crystallization conditions,[90] the two states are equally probable and thus equal in free energy. This scenario is akin to classic conformational selection,[91] highlighting that both conformations are thermally accessible in the absence of CODH substrates. Given the observation of substrate access tunnels in both states, while only state B is thought to be competent for chemistry,[90] allosteric regulation for the CODH structure may be one mechanism for guiding metabolism through this multifunctional enzyme.[92] Taken together, between the ADK and CODH examples, the Multi-Map colvars capture two of the most universal mechanisms of allosteric interactions.

IV. CONCLUSION

The use of volumetric maps as sources of external potentials in MD simulations has allowed the development of many enhanced sampling methods such as grid-steered MD,[51] MDFF,[31,48] and atom resolved Brownian dynamics.[93] The most recent addition to these methods is the Multi-Map variable,[37] which is applied here to monitor transitions in the protein structure based on electron density maps from cryo-EM or crystallography. Transformations between two protein states are successfully sampled for different protein sizes and types of conformational changes, suggesting that the method is generalizable to a series of cryo-EM maps akin to the outcome of the manifold-based cryo-EM data analysis.[29,54] In addition to the simultaneous utility of the Multi-Map colvar toward non-equilibrium work and free energy estimation, we here demonstrate its usefulness in the refinement of map–structure correlation coefficients and model-quality without requiring high-quality search structures. The Multi-Map formulation offers a statistical mechanical description for the resolvability of a map, addressing a point of concern in cryo-EM modeling.

Previous work demonstrated that the Multi-Map variable can characterize shape changes in supra-molecular aggregates such as biological membranes or confined-water pockets.[37,38] The results shown here also open the door to its application together with a number of geometric or alchemical free energy methods focused on protein conformational cycles.[94] Furthermore, in the specific application of structure refinement, data-driven approaches such as MELD (Modeling Employing Limited Data) [61] or meta-inference[95] can readily employ data simulated using the Multi-Map variable as a source of coarse-grained information for computing the Bayesian priors. From a biophysical standpoint, two of the most universal allosteric pathways, namely, induced fit and conformational selection, were here successfully investigated. In this work, we used simulated density maps; however, these protocols are shown to be equally applicable to the experimentally determined density maps. Through the use of simulated maps, we were able to determine map resolution's effect on energy cost of flexible fitting. The free energy studies were found to be most efficient using low-resolution maps, and the estimates are comparable to those determined from high-resolution structures.

SUPPLEMENTARY MATERIAL

See the supplementary material for details regarding simulation parameters and calculations of collective variables other than the Multi-Map variable, as well as equilibrium, SMD, and BEUS simulation analysis figures.

ACKNOWLEDGMENTS

We acknowledge start-up funds from the SMS and CASD at Arizona State University. This research used resources and is authored, in part, by the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. A.S. acknowledges the CAREER award from NSF Grant Nos. MCB-1942763 and NIH/R01GM095583. J.W.V. acknowledges the support from the National Science Foundation Graduate Research Fellowship under Grant No. 2020298734. The authors acknowledge Research Computing at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin and the Arizona State University for providing HPC, visualization, database, or grid resources that have contributed to the research results reported within this paper.

APPENDIX A: ESTIMATING ENDPOINTS FOR THE MULTI-MAP COLVAR FOR DIFFERENT RESOLUTIONS

The Multi-Map collective variable used here is defined in compact form as ζ_{AB} [Eq. (2)] by using two three-dimensional volumetric maps $\phi_A(\mathbf{r})$ and $\phi_B(\mathbf{r})$, which are synthetically generated but are treated otherwise as experimental data. Based only on the atomic density maps, it is thus possible to estimate what the colvar range should be. Given an atomic configuration $\mathbf{R}$, a synthetic map $\phi_{\mathbf{R}}(\mathbf{r})$ may be generated and its cross-correlation with $\phi_A(\mathbf{r})$ defined as

$$CC(\mathbf{R},A) = \frac{1}{|\phi_{\mathbf{R}}||\phi_A|}\int_{\mathbf{r}} \phi_{\mathbf{R}}(\mathbf{r})\phi_A(\mathbf{r})d\mathbf{r} \tag{A1}$$

and similarly for $CC(\mathbf{R}, B)$. Inserting Eq. (A1) into Eq. (2), we arrive at

$$\zeta_{AB}(\mathbf{R}) \simeq \Phi(\mathbf{R})^{-1}[CC(\mathbf{R},B) - CC(\mathbf{R},A)], \tag{A2}$$

where the approximation lies in assuming that the density map $\phi_{\mathbf{R}}$ is used *in lieu* of the precise atomic coordinates $\mathbf{R}$. In addition, $\Phi(\mathbf{R}) = \sum_i w_i \phi(\mathbf{r}_i)$ is the term of the Multi-Map variable evaluated at the coordinates $\mathbf{R}$ that best fit the set of local map ϕ. It is essential at this stage to recognize that correlation coefficient values fall in the range $(0, 1)$ due to non-negative densities. For instance, if a 0 Å resolution map were to exist, the densities would be delta functions with values equal to the atomic weight (w_i), meaning that $\Phi(\mathbf{R})^{-1} = (\sum_i w_i)^{-1}$. For other resolutions, $\Phi(\mathbf{R})^{-1}$ will similarly be a scaling factor related to the total atomic weight and will be invariant for well-fitted maps with equivalent resolution. Stated concretely for this special mass-conserving transition between states A to B,

$$\zeta_A(\mathbf{R}_A) = \zeta_B(\mathbf{R}_B). \tag{A3}$$

Assuming that a map is a Gaussian mixture model, the equality holds if both maps have homogeneous and equivalent resolutions. If the resolutions are not equivalent, then a scaling factor is needed to maintain the equality, as outlined below. However, a more general boundary condition is

$$CC(A, A) = CC(B, B) = 1. \tag{A4}$$

Combining Eqs. (A2) and (A3),

$$\zeta_{AB}(\mathbf{R}_A) = \Phi(\mathbf{R}_A)^{-1}[CC(A,B) - CC(A,A)], \tag{A5}$$

$$\zeta_{AB}(\mathbf{R}_B) = \Phi(\mathbf{R}_B)^{-1}[CC(B,B) - CC(B,A)], \tag{A6}$$

and therefore,

$$\zeta_{AB}(\mathbf{R}_A) = -\zeta_{AB}(\mathbf{R}_B) \tag{A7}$$

because $CC(A, B) = CC(B, A)$. In practice, Eq. (A7) is limited by discretization errors because both maps are interpolated onto a grid. However, by comparing the numerical and theoretical values $\zeta_{AB}(\mathbf{R}_A) = \Phi_t^{-1}[CC(A,B) - CC(A,A)]$ determined from Eqs. (2) and (A2) in Table S4 within Sec. S3 for a range of resolutions, we find that a 1 Å grid spacing has acceptable numerical error. As a consequence of Eq. (A7), the target value for ζ_{AB} can be estimated, knowing only one endpoint from the transition for two maps with equal resolution. Importantly, all other atomic configurations other than $\mathbf{R}_A$ or $\mathbf{R}_B$ will generate ζ_{AB} values whose magnitude is less than $|\zeta_{AB}(\mathbf{R}_A)| = |\zeta_{AB}(\mathbf{R}_B)|$, as the correlation coefficients are bounded and the density weight is conserved [Eq. (A3)].

If two maps have unequal resolutions with comparable local densities, the range for ζ_{AB} will depend on the ratio between $\Phi_A(\mathbf{R}_A)$ and $\Phi_B(\mathbf{R}_B)$ like

$$\zeta_{AB}(\mathbf{R}_A) = -\frac{\frac{\exp\left(\sigma_B^{-2}\right)}{\sigma_B}}{\frac{\exp\left(\sigma_A^{-2}\right)}{\sigma_A}}\zeta_{AB}(\mathbf{R}_B), \tag{A8}$$

where σ_A and σ_B are the resolutions of maps A and B. The local density is calculated as the sum of atom weights divided by the map volume, which can be computed in VMD. Thus, even when the resolutions of the contributing maps are different, but their local densities are similar, we can still determine $\zeta_{AB}(\mathbf{R}_B)$ with the knowledge of only one high-quality structure across a series of maps. However, when the resolution of a map is nonuniform and the local densities do not match between the states A and B, the assumptions of Eqs. (A7) and (A8) fail. In these cases, structural information on both $\mathbf{R}_A$ and $\mathbf{R}_B$ is needed *a priori* to determine the endpoint values of ζ_{AB} for subsequent application in equilibrium and non-equilibrium MD or enhanced sampling simulations.

These considerations are non-trivially generalizable to capture conformational changes across the entire series of K-maps, which are considered a sum of two-map transformations. Thus, by knowing the structure R at only one endpoint, the Multi-Map colvar allows, in principle, the construction of ensembles and simultaneous real-space refinement for each of the K maps contributing to the colvar. In practice, the variation of the local resolution within a cryo-EM map may prevent the application simple scaling rules to all the maps. Nonetheless, the nominal resolution of majority of the maps coming from conformational analysis with EM has highly comparable resolution as seen in the ribosome,[28] RyR1 receptor,[29] and recently in spike protein.[96]

APPENDIX B: IMPLEMENTATION AND AVAILABILITY OF THE MULTI-MAP COLLECTIVE VARIABLE

The derivation and implementation of the Multi-Map variable are documented in Ref. 37. The implementation leverages recent improvements to the GridForces[51] and Colvars[60] modules, both of which are freely available in the most recent version of NAMD.[47] Up-to-date documentation and input file fragments for several use cases are available at https://colvars.github.io/colvars-refman-namd/colvars-refman-namd.html#sec:cvc_multimap.

A NAMD configuration file fragment for invoking the two-state colvar (i.e. where the coefficients of the Multi-Map variable are -1 and +1, respectively) is provided below:

```
# Load the two electron density maps
set GRIDFILE [list stateA.dx stateB.dx]
mgridForce on
for {set i 0} {$i <[llength $GRIDFILE]} {incr i} {
    mgridForceFile $i gridpdb.pdb ; # Flag coupled atoms
    mgridForceCol $i 0; #Couple to the map if 0 ! = 0
    mgridForceChargeCol $i B; #Weight based on value in B
    mgridForcePotFile $i [lindex $GRIDFILE $i];
    mgridForceScale $i 0 0 0; #Maps do not contribute bias themselves,
    #except through the Multi-Map colvar, which is defined below.
}
# Initialize Zeta_AB colvar
    colvars on
    cv config "
    colvar {
       name zeta_AB
       mapTotal {
        #Maps are 0-indexed, so map 0 is state A.
          mapName 0
          componentCoeff - 1
       }
       mapTotal {
        #Map 1 is state B
           mapName 1
           componentCoeff 1
       }
}
"
## End code ##
```

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material. The underlying trajectories and analysis scripts will be released using the Constellation Portal at the OLCF, https://doi.ccs.ornl.gov/ui/doi/108, with doi 10.13139/OLCF/1674982.

REFERENCES

[1] S. Subramaniam, "The cryo-EM revolution: Fueling the next phase," IUCrJ **6**, 1–2 (2019).

[2] K. M. Yip, N. Fischer, E. Paknia, A. Chari, and H. Stark, "Breaking the next Cryo-EM resolution barrier—Atomic resolution determination of proteins!," bioRxiv:2020.05.21.106740 (2020).

[3] E. H. Egelman, "The current revolution in cryo-EM," Biophys. J. **110**, 1008–1012 (2016).

[4] M. A. Herzik, M. Wu, and G. C. Lander, "High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM," Nat. Commun. **10**, 1032 (2019).

[5] A. H. de la Peña, E. A. Goodall, S. N. Gates, G. C. Lander, and A. Martin, "Substrate-engaged 26S proteosome structures reveal mechanisms for ATP-hydrolysis–driven translocation," Science **362**, eaav0725 (2018).

[6] X. Zhan, C. Yan, X. Zhang, J. Lei, and Y. Shi, "Structure of a human catalytic step I spliceosome," Science **359**, 537–545 (2018).

[7] S. Stolzenberg, M. Michino, M. V. LeVine, H. Weinstein, and L. Shi, "Computational approaches to detect allosteric pathways in transmembrane molecular machines," Biochim. Biophys. Acta, Biomembr. **1858**, 1652–1662 (2016).

[8] Y. Karami, T. Bitard-Feildel, E. Laine, and A. Carbone, "'Infostery' analysis of short molecular dynamics simulations identifies highly sensitive residues and predicts deleterious mutations," Sci. Rep. **8**, 16126 (2018).

[9] S. Ikemura, H. Yasuda, S. Matsumoto, M. Kamada, J. Hamamoto, K. Masuzawa, K. Kobayashi, T. Manabe, D. Arai, I. Nakachi, I. Kawada, K. Ishioka, M. Nakamura, H. Namkoong, K. Naoki, F. Ono, M. Araki, R. Kanada, B. Ma, Y. Hayashi, S. Mimaki, K. Yoh, S. S. Kobayashi, T. Kohno, Y. Okuno, K. Goto, K. Tsuchihara, and K. Soejima, "Molecular dynamics simulation-guided drug sensitivity prediction for lung cancer with rare EGFR mutations," Proc. Natl. Acad. Sci. U. S. A. **116**, 10025–10030 (2019).

[10] P. E. M. Lopes, O. Guvench, and A. D. MacKerell, "Current status of protein force fields for molecular dynamics simulations," in *Methods in Molecular Biology*, edited by A. Kukol (Humana Press, New York, NY, USA, 2015), pp. 47–71.

[11] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Interdisciplinary Applied Mathematics Vol. 21 (Springer New York, NY, 2010).

[12] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," Phys. Rev. Lett. **55**, 2471–2474 (1985).

[13] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, "*Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals," J. Chem. Phys. **114**, 9758–9763 (2001).

[14] M. E. Tuckerman, "*Ab initio* molecular dynamics: Basic concepts, current trends and novel applications," J. Phys.: Condens. Matter **14**, R1297–R1355 (2002).

[15] Free Energy Calculations, Springer Series in Chemical Physics Vol. 86, edited by C. Chipot and A. Pohorille (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).

[16] W. L. Jorgensen and C. Ravimohan, "Monte Carlo simulation of differences in free energies of hydration," J. Chem. Phys. **83**, 3050–3054 (1985).

[17] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**, 187–199 (1977).

[18] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," J. Comput. Chem. **13**, 1011–1021 (1992).

[19] H. Grubmüller, "Predicting slow structural transitions in macromolecular systems: Conformational flooding," Phys. Rev. E **52**, 2893–2906 (1995).

[20] A. Laio and M. Parrinello, "Escaping free-energy minima," Proc. Natl. Acad. Sci. U. S. A. **99**, 12562–12566 (2002).

[21] A. C. Pan, D. Sezer, and B. Roux, "Finding transition pathways using the string method with swarms of trajectories," J. Phys. Chem. B **112**, 3432–3440 (2008).

[22] Y. Miao, V. A. Feher, and J. A. McCammon, "Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation," J. Chem. Theory Comput. **11**, 3584–3595 (2015).

[23] C. R. Schwantes and V. S. Pande, "Improvements in markov state model construction reveal many non-native interactions in the folding of NTL9," J. Chem. Theory Comput. **9**, 2000–2009 (2013).

[24] D. Spetzler, J. York, D. Daniel, R. Fromme, D. Lowry, and W. Frasch, "Microsecond time scale rotation measurements of single F_1-ATPase molecules," Biochemistry **45**, 3117–3124 (2006).

[25] J. L. Martin, R. Ishmukhametov, D. Spetzler, T. Hornung, and W. D. Frasch, "Elastic coupling power stroke mechanism of the F_1-ATPase molecular motor," Proc. Natl. Acad. Sci. U. S. A. **115**, 5750–5755 (2018).

[26] R. Laskowski, J. A. Rullmann, M. MacArthur, R. Kaptein, and J. Thornton, "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR," J. Biomol. NMR **8**, 477 (1996).

[27] B. Roux and S. M. Islam, "Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron–electron resonance spectroscopy," J. Phys. Chem. B **117**, 4733–4739 (2013).

[28] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinizadeh, H. Y. Liao, J. Pallesen, G. Sharma, V. A. Stupina, A. E. Simon, J. D. Dinman, J. Frank, and A. Ourmazd, "Trajectories of the ribosome as a Brownian nanomachine," Proc. Natl. Acad. Sci. U. S. A. **111**, 17492–17497 (2014).

[29] A. Dashti, M. S. Shekhar, D. B. Hail, G. Mashayekhi, P. Schwander, A. des Georges, J. Frank, A. Singharoy, and A. Ourmazd, "Functional pathways of biomolecules retrieved from single-particle snapshots," bioRxiv:291922 (2019).

[30] F. Tama, O. Miyashita, and C. L. Brooks III, "Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM," J. Struct. Biol. **147**, 315–326 (2004).

[31] A. Singharoy, I. Teo, R. McGreevy, J. E. Stone, J. Zhao, and K. Schulten, "Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps," eLife **5**, e16105 (2016).

[32] A. Perez, J. L. MacCallum, and K. A. Dill, "Accelerating molecular simulations of proteins using Bayesian inference on weak information," Proc. Natl. Acad. Sci. U. S. A. **112**, 11846–11851 (2015).

[33] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, and P. D. Adams, "Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in phenix," Acta Crystallogr., Sect. D: Struct. Biol. **75**, 861–877 (2019).

[34] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, "Overview of the CCP4 suite and current developments," Acta Crystallogr., Sect. D: Struct. Biol. **67**, 235–242 (2011).

[35] L. Eshun-Wilson, R. Zhang, D. Portran, M. V. Nachury, D. B. Toso, T. Löhr, M. Vendruscolo, M. Bonomi, J. S. Fraser, and E. Nogales, "Effects of α-tubulin acetylation on microtubule structure and stability," Proc. Natl. Acad. Sci. U. S. A. **116**, 10366–10371 (2019).

[36] M. Shekhar, G. Terashi, C. Gupta, G. Debussche, N. J. Sisco, J. Nguyen, J. Zook, J. Vant, D. Sarkar, P. Fromme, W. D. Van Horn, K. Dill, D. Kihara, E. Tajkhorshid, A. Perez, and A. Singharoy, "CryoFold: *Ab-initio* structure determination from electron density maps using molecular dynamics," bioRxiv:687087 (2019).

[37] G. Fiorin, F. Marinelli, and J. D. Faraldo-Gómez, "Direct derivation of free energies of membrane deformation and other solvent density variations from enhanced sampling molecular dynamics," J. Comput. Chem. **41**, 449–459 (2020).

[38] W. Zhou, G. Fiorin, C. Anselmi, H. A. Karimi-Varzaneh, H. Poblete, L. R. Forrest, and J. D. Faraldo-Gómez, "Large-scale state-dependent membrane remodeling by a transporter protein," eLife **8**, e50576 (2019).

[39] G. Pintilie, K. Zhang, Z. Su, S. Li, M. F. Schmid, and W. Chiu, "Measurement of atom resolvability in cryo-EM maps with Q-scores," Nat. Methods **17**, 328–334 (2020).

[40] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," J. Mol. Graphics **14**, 33–38 (1996).

[41] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," Nat. Methods **14**, 71–73 (2017).

[42] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, and M. L. Pastor, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926 (1983).

[43] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," J. Comput. Chem. **13**, 952–962 (1992).

[44] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," J. Chem. Phys. **101**, 4177–4189 (1994).

[45] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, "Constant pressure molecular dynamics simulation: The Langevin piston method," J. Chem. Phys. **103**, 4613–4621 (1995).

[46] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**, 8577 (1995).

[47] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," J. Comput. Chem. **26**, 1781–1802 (2005).

[48] L. G. Trabuco, E. Villa, K. Mitra, J. Frank, and K. Schulten, "Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics," Structure **16**, 673–683 (2008).

[49] L. G. Trabuco, E. Villa, E. Schreiner, C. B. Harrison, and K. Schulten, "Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography," Methods **49**, 174–180 (2009).

[50] R. McGreevy, I. Teo, A. Singharoy, and K. Schulten, "Advances in the molecular dynamics flexible fitting method for cryo-EM modeling," Methods **100**, 50–60 (2016).

[51] D. B. Wells, V. Abramkina, and A. Aksimentiev, "Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics," J. Chem. Phys. **127**, 125101 (2007).

[52] V. Pinto, J. Araújo, R. Silva, G. da Costa, J. Cruz, M. De A. Neto, J. Campos, C. Santos, F. Leite, and M. Junior, "In silico study to identify new antituberculosis molecules from natural sources by hierarchical virtual screening and molecular dynamics simulations," Pharmaceuticals **12**, 36 (2019).

[53] A. Das, M. Gur, M. H. Cheng, S. Jo, I. Bahar, and B. Roux, "Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model," PLoS Comput. Biol. **10**, e1003521 (2014).

[54] J. Frank and A. Ourmazd, "Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM," Methods **100**, 61–67 (2016).

[55] E. Eyal, L.-W. Yang, and I. Bahar, "Anisotropic network model: Systematic evaluation and a new web interface," Bioinformatics **22**, 2619–2627 (2006).

[56] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution," Proc. Natl. Acad. Sci. U. S. A. **110**, 18916–18921 (2013).

[57] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," Nucleic Acids Res. **28**, 235–242 (2000).

[58] B. Frenz, A. C. Walls, E. H. Egelman, D. Veesler, and F. DiMaio, "RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps," Nat. Methods **14**, 797–800 (2017).

[59] B. A. Barad, N. Echols, R. Y.-R. Wang, Y. Cheng, F. DiMaio, P. D. Adams, and J. S. Fraser, "EMRinger: Side chain–directed model and map validation for 3D cryo-electron microscopy," Nat. Methods **12**, 943–946 (2015).

[60] G. Fiorin, M. L. Klein, and J. Hénin, "Using collective variables to drive molecular dynamics simulations," Mol. Phys. **111**, 3345–3362 (2013).

[61] J. L. MacCallum, A. Perez, and K. A. Dill, "Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference," Proc. Natl. Acad. Sci. U. S. A. **112**, 6985–6990 (2015).

[62] B. W. J. Irwin and D. J. Huggins, "Estimating atomic contributions to hydration and binding using free energy perturbation," J. Chem. Theory Comput. **14**, 3218–3227 (2018).

[63] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, "The adaptive biasing force method: Everything you always wanted to know but were afraid to ask," J. Phys. Chem. B **119**, 1129–1151 (2015).

[64] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," Phys. Rev. Lett. **100**, 020603 (2008).

[65] M. Moradi and E. Tajkhorshid, "Computational recipe for efficient description of large-scale conformational changes in biomolecular systems." J. Chem. Theory Comput. **10**, 2866–2880 (2014).

[66] Y. Sugita, A. Kitao, and Y. Okamoto, "Multidimensional replica-exchange method for free-energy calculations," J. Chem. Phys. **113**, 6042–6051 (2000).

[67] J. V. Vermaas, R. Kont, G. T. Beckham, M. F. Crowley, M. Gudmundsson, M. Sandgren, J. Ståhlberg, P. Väljamäe, and B. C. Knott, "The dissociation mechanism of processive cellulases," Proc. Natl. Acad. Sci. U. S. A. **116**, 23061–23067 (2019).

[68] C. Jarzynski, "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale," Annu. Rev. Condens. Matter Phys. **2**, 329–351 (2011).

[69] A. L. Ferguson, "BayesWHAM: A Bayesian approach for free energy estimation, reweighting, and uncertainty quantification in the weighted histogram analysis method," J. Comput. Chem. **38**, 1583–1605 (2017).

[70] M. Habeck, "Bayesian estimation of free energies from equilibrium simulations," Phys. Rev. Lett. **109**, 100601 (2012).

[71] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," J. Chem. Phys. **129**, 124105 (2008).

[72] B. Isralewitz, S. Izrailev, and K. Schulten, "Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations," Biophys. J. **73**, 2972–2979 (1997).

[73] J. Rydzewski, R. Jakubowski, W. Nowak, and H. Grubmüller, "Kinetics of huperzine A dissociation from acetylcholinesterase via multiple unbinding pathways," J. Chem. Theory Comput. **14**, 2843–2851 (2018).

[74] J. Zook, M. Shekhar, D. Hansen, C. Conrad, T. Grant, C. Gupta, T. White, A. Barty, S. Basu, Y. Zhao, N. Zatsepin, A. Ishchenko, A. Batyuk, C. Gati, C. Li, L. Galli, J. Coe, M. Hunter, M. Liang, U. Weierstall, G. Nelson, D. James, B. Stauch, F. Craciunescu, D. Thifault, W. Liu, V. Cherezov, A. Singharoy, and P. Fromme, "XFEL and NMR structures of Francisella lipoprotein reveal conformational space of drug target against tularemia," Structure **28**, 540, e3–547 (2020).

[75] M. G. Rossmann and D. M. Blow, "The detection of sub-units within the crystallographic asymmetric unit," Acta Crystallogr. **15**, 24–31 (1962).

[76] F. Fabiola and M. S. Chapman, "Fitting of high-resolution structures into electron microscopy reconstruction images," Structure **13**, 389–400 (2005).

[77] M. Igaev, C. Kutzner, L. V. Bock, A. C. Vaiana, and H. Grubmüller, "Automated cryo-EM structure refinement using correlation-driven molecular dynamics," eLife **8**, e43542 (2019).

[78] F. DiMaio, Y. Song, X. Li, M. J. Brunner, C. Xu, V. Conticello, E. Egelman, T. C. Marlovits, Y. Cheng, and D. Baker, "Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement," Nat. Methods **12**, 361–365 (2015).

[79] C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, and D. C. Richardson, "MolProbity: More and better reference data for improved all-atom structure validation," Protein Sci. **27**, 293–315 (2018).

[80] K. Arora and C. L. Brooks, "Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism," Proc. Natl. Acad. Sci. U. S. A. **104**, 18496–18501 (2007).

[81] Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, and A. Kidera, "Minimum free energy path of ligand-induced transition in adenylate kinase," PLoS Comput. Biol. **8**, e1002555 (2012).

[82] D. Li, M. S. Liu, and B. Ji, "Mapping the dynamics landscape of conformational transitions in enzyme: The adenylate kinase case," Biophys. J. **109**, 647–660 (2015).

[83] H. Y. Aviram, M. Pirchi, H. Mazal, Y. Barak, I. Riven, and G. Haran, "Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions," Proc. Natl. Acad. Sci. U. S. A. **115**, 3243–3248 (2018).

[84] U. Olsson and M. Wolf-Watz, "Overlap between folding and functional energy landscapes for adenylate kinase conformational change," Nat. Commun. **1**, 111 (2010).

[85] H. D. Song and F. Zhu, "Conformational dynamics of a ligand-free adenylate kinase," PLoS One **8**, e68023 (2013).

[86] E. F. Pai, W. Sachsenheimer, R. H. Schirmer, and G. E. Schulz, "Substrate positions and induced-fit in crystalline adenylate kinase," J. Mol. Biol. **114**, 37–45 (1977).

[87] M. Kovermann, C. Grundström, A. E. Sauer-Eriksson, U. H. Sauer, and M. Wolf-Watz, "Structural basis for ligand binding to an enzyme by a conformational selection pathway," Proc. Natl. Acad. Sci. U. S. A. **114**, 6298–6303 (2017).

[88] J. Zook, G. Mo, N. J. Sisco, F. M. Craciunescu, D. T. Hansen, B. Baravati, B. R. Cherry, K. Sykes, R. Wachter, W. D. Van Horn, and P. Fromme, "NMR structure of Francisella tularensis virulence determinant reveals structural homology to bet v1 allergen proteins," Structure **23**, 1116–1122 (2015).

[89] P.-h. Wang, M. Bruschi, L. De Gioia, and J. Blumberger, "Uncovering a dynamically formed substrate access tunnel in carbon monoxide dehydrogenase/acetyl-CoA synthase," J. Am. Chem. Soc. **135**, 9493–9502 (2013).

[90] C. Darnault, A. Volbeda, E. J. Kim, P. Legrand, X. Vernède, P. A. Lindahl, and J. C. Fontecilla-Camps, "Ni-Zn-[Fe_4-S_4] and Ni-Ni-[Fe_4-S_4] clusters in closed and open α subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase," Nat. Struct. Mol. Biol. **10**, 271–279 (2003).

[91] F. Paul and T. R. Weikl, "How to distinguish conformational selection and induced fit based on chemical relaxation rates," PLoS Comput. Biol. **12**, e1005067 (2016).

[92] M. Can, F. A. Armstrong, and S. W. Ragsdale, "Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase," Chem. Rev. **114**, 4149–4174 (2014).

[93] J. Comer and A. Aksimentiev, "Predicting the DNA sequence dependence of nanopore ion current using atomic-resolution Brownian dynamics," J. Phys. Chem. C **116**, 3376–3393 (2012).

[94] A. Singharoy, C. Chipot, M. Moradi, and K. Schulten, "Chemomechanical coupling in hexameric protein–protein interfaces harnesses energy within V-type ATPases," J. Am. Chem. Soc. **139**, 293–310 (2017).

[95] M. Bonomi, C. Camilloni, A. Cavalli, and M. Vendruscolo, "Metainference: A Bayesian inference method for heterogeneous systems," Sci. Adv. **2**, e1501177 (2016).

[96] R. Melero, C. O. S. Sorzano, B. Foster, J.-L. Vilas, M. Martínez, R. Marabini, E. Ramírez-Aportela, R. Sanchez-Garcia, D. Herreros, L. del Caño, P. Losana, Y. C. Fonseca-Reyna, P. Conesa, D. Wrapp, P. Chacon, J. S. McLellan, H. D. Tagare, and J.-M. Carazo, "Continuous flexibility analysis of SARS-CoV-2 Spike prefusion structures," bioRxiv:2020.07.08.191072 (2020).

[97] C. W. Müller and G. E. Schulz, "Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap_5A refined at 1.9 Å resolution: A model for a catalytic transition state," J. Mol. Biol. **224**, 159–177 (1992).

[98] C. Müller, G. Schlauderer, J. Reinstein, and G. Schulz, "Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding," Structure **4**, 147–156 (1996).