Multi-level free energy simulation with a staged transformation approach

View Online   Export Citation   CrossMark

Shingo Ito[1] and Qiang Cui[2,a] (ID)

AFFILIATIONS

[1] Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA

[2] Departments of Chemistry, Physics and Biomedical Engineering, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA

Note: This paper is part of the JCP Special Topic on Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force Fields, and Applications.
[a] Author to whom correspondence should be addressed: qiangcui@bu.edu. **Tel.:** (+1)-617-353-6189

ABSTRACT

Combining multiple levels of theory in free energy simulations to balance computational accuracy and efficiency is a promising approach for studying processes in the condensed phase. While the basic idea has been proposed and explored for quite some time, it remains challenging to achieve convergence for such multi-level free energy simulations as it requires a favorable distribution overlap between different levels of theory. Previous efforts focused on improving the distribution overlap by either altering the low-level of theory for the specific system of interest or ignoring certain degrees of freedom. Here, we propose an alternative strategy that first identifies the degrees of freedom that lead to gaps in the distributions of different levels of theory and then treats them separately with either constraints or restraints or by introducing an intermediate model that better connects the low and high levels of theory. As a result, the conversion from the low level to the high level model is done in a staged fashion that ensures a favorable distribution overlap along the way. Free energy components associated with different steps are mostly evaluated explicitly, and thus, the final result can be meaningfully compared to the rigorous free energy difference between the two levels of theory with limited and well-defined approximations. The additional free energy component calculations involve simulations at the low level of theory and therefore do not incur high computational costs. The approach is illustrated with two simple but non-trivial solution examples, and factors that dictate the reliability of the result are discussed.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0012494

I. INTRODUCTION

Hybrid quantum mechanical/molecular mechanical (QM/MM) methods[1–9] have become an established technique in computational chemistry, biochemistry, and biophysics as well as materials science.[7,9–16] One of the remaining challenges for predictive applications for condensed phase systems is to compute reliable free energy changes. As discussed in numerous studies,[17–20] this requires, in principle, an accurate QM/MM potential function and adequate sampling, which is difficult to accomplish for most condensed phase studies at this stage. Therefore, a potentially attractive alternative is to conduct multi-level free energy simulations[20–29] in which a low-level QM/MM potential (e.g., using a semi-empirical QM method) is used to conduct extensive sampling, while a higher-level QM/MM

potential (e.g., with a density functional theory or a correlated wave-function method as QM) is used to improve the free energy estimate; if the process of interest is not reactive in nature, the low-level method can be a classical MM model as well. A commonly discussed free energy cycle is shown in Fig. 1; due to the state function nature of free energy, $\Delta G_{A \to B}^H = \Delta G_{A \to B}^L + \Delta G_B^{L \to H} - \Delta G_A^{L \to H}$, where "L" and "H" indicate the low and high levels of theory, respectively. For the quantities on the rhs of the equation, $\Delta G_{A \to B}^L$ is obtained with a low-level of theory and therefore can be computed with extensive sampling using, for example, umbrella sampling[30] or its more modern variants.[31] The "vertical" components, $\Delta G_{A/B}^{L \to H}$, are needed only for the end-states, A/B; for the multi-level free energy simulation to be computationally worthwhile, calculations of $\Delta G_{A/B}^{L \to H}$ should rely mostly (if not exclusively) on sampling at the low-level, and high-level
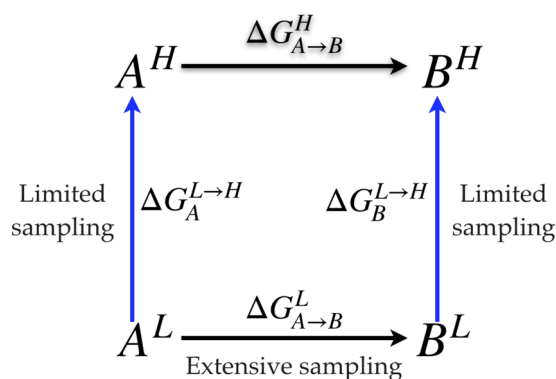
FIG. 1. A thermodynamic cycle that uses the low-level (L) method as a reference potential to indirectly compute the free energy difference between states A and B at the high (H) level. In this work, we focus on a single vertical leg and discuss the computation of $\Delta G^{L \to H}$.

single point energy calculations are done only at a limited number of uncorrelated snapshots. The most naïve estimator, for example, is the one-step free energy perturbation, $\Delta G_{A/B}^{L \to H} = -k_B T \ln \langle e^{-\beta \Delta U_{A/B}^{L \to H}} \rangle_L$, which, in principle, requires sampling only at the low level.

While this is an intuitive idea and has been explored for quite a few times in previous studies in slightly different forms,[20–29] reaching converged results, in practice, for $\Delta G_{A/B}^{L \to H}$ based on low-level sampling alone has been difficult, except for relatively simple or structurally rigid systems. This is not unexpected since it has been well-known in the free energy simulation literature that a single step perturbation gives reasonable results only if the two end-states (in this case, low- and high-level potential energies) are sufficiently similar such that the standard deviation of energy difference (in our case, $\Delta U_{A/B}^{L \to H}$) is on the order of several $k_B T$.[32–35] More quantitative metrics for the distribution overlap between the two end-states have been developed by several authors,[32,33,36,37] but most of them are not straightforward to use for the case of multi-level free energy simulations in which high-level sampling is to be avoided as much as possible. Accordingly, based on the work of Kofke and co-workers,[38–40] Boresch and Woodcock developed an approximate version[41] of the Π score[38,39] to estimate the degree of distribution overlap between two levels of theory; a value of >0.5 was recommended as the indicator of the minimal overlap for the meaningful one-step free energy estimate. It was found that the approximate Π score correlates reasonably well with the width of the $\Delta U^{L \to H}$ distribution, which can also be an instructive factor for monitoring the convergence of one-step free energy estimates.[33,34,41]

The test of a range of small molecules in both gas phase and solution using CHARMM22/36 as the low-level and DFTB3[42,43] as the high-level indicated that the overlap Π score is favorable for rigid molecules.[44] For more flexible molecules, such as a serine amino acid in the gas phase, the overlap is poor with a negative Π score.[45] To improve the convergence behavior, several strategies have been developed. If the main difference between the two levels of theory concerns the bonded degrees of freedom, one approach is to conduct non-equilibrium free energy simulations by rapidly

switching the potential function; since the bonded degrees of freedom (especially bond stretch and angle bending) tend to be highly localized, rapid switching at the picosecond time scale appeared effective, although many trajectories (on the order of thousands or more) are usually required for stringent numerical convergence, which is a well-documented feature for non-equilibrium free energy simulations[46] based on the Jarzynski equality[47] or the fluctuation theorem.[48]

When more general degrees of freedom are responsible for the deviation between different levels of theory, it seems necessary to modify the low-level of theory such that the distribution overlap with the high-level method is improved. This has been done by several authors in the broader context of free energy simulations. For example, Rossi and Truhlar[49] pioneered reaction-specific-parameters to improve the semi-empirical QM method for the specific system of interest; this has been adapted in the framework of the string method with force matching by Zhou et al.[50] for effective reaction free energy simulations. Similarly, Shen andYang[51] reported an approach in which a neural network is used to iteratively improve the low-level QM method in reaction free energy simulations. Plotnikov et al. developed "para-dynamics,"[52] which iteratively improves the reference potential in the empirical valence bond (EVB) framework using ab initio QM single point energies. Specifically, in the context of multi-level free energy simulation based on the thermodynamic cycle in Fig. 1, Heimdal and Ryde advocated for the parameterization of better force fields as the reference approach.[19] Similarly, Woodcock and co-workers[45,53] as well as York and co-workers[54,55] proposed to use force-matching to improve the low-level potential based on single-point high-level force calculations; for relatively rigid small molecules in solution or a protein active site, encouraging results have been reported. For rigid molecules, another possibility is to compute only the QM/MM interaction energies,[56] although this is clearly an approximation and may lead to considerable errors as the flexibility of molecules increases.[57]

Despite this progress, it is worthwhile exploring alternative strategies for connecting different levels of theory in quantitative free energy simulations. For example, while the non-equilibrium approach can benefit from "embarrassingly parallel" simulations,[53] conducting thousands of picosecond scale trajectories that involve explicitly mixing low- and high-levels of potentials is computationally demanding when the QM region is large; moreover, conducting many fast non-equilibrium switching simulations likely entangles the free energy difference due to the stiff bonded degrees of freedom with that arising from incomplete sampling of other degrees of freedom. Similarly, improving a low-level method with force matching or even a neural network may not be straightforward for large systems.

Motivated by these considerations, we explore an approach in which the low-level theory is not modified explicitly; instead, the transformation from the low-level to the high-level potential function is done in a staged fashion such that a favorable distribution overlap is always maintained. Although the transformation is carried out in multiple stages and thus involves more computations, most of these computations are done at the low-level of theory, and therefore, the approach still offers a significant gain in computational efficiency compared to brute-force free energy simulations that explicitly mix low and high level potential functions in multiple windows.

In the following, we first describe the fundamental consideration behind our development and then present the staged transformation approach. Next, as a proof of concept, we study two small but non-trivial molecules in solution and discuss technical choices that are likely important to the effectiveness of our specific thermodynamic path. We end with a few concluding remarks, including our perspective regarding future developments.

II. METHOD

In the following discussion, we focus on a single vertical leg of the thermodynamic cycle in Fig. 1. To simplify nomenclature, we make the assumption that the A/B conversion depicted in the generic thermodynamic cycle in Fig. 1 involves the entire QM region only, which will also be simply referred to as the "solute."

A. Fundamental considerations

As shown schematically in Fig. 2, different strategies can be used to improve the distribution overlap between L and H levels of theory [panel (a)]. The most natural approach, as mentioned above in the Introduction, is to modify the low level of theory [$L \to L'$, panel (b)]. Since this may not be straightforward, one could consider other alternatives. One possibility is to artificially broaden the distribution at the low level of theory (i.e., sample P'_L rather than P_L with, for example, a higher temperature or Tsallis statistics[58]) so as to develop the distribution overlap with H; to complete the thermodynamic cycle, however, one has to either reweight P'_L or compute the free energy difference between the original and broadened distributions. In fact, panel (c) can be considered as a special case for panel (d) in which an intermediate distribution P_M is introduced to bridge the otherwise separated P_L and P_H; P_M can be sampled using a different level of theory, using L in a different thermodynamic state, or using L in the presence of additional constraints/restraints (see below). The fundamental philosophy is that instead of modifying L explicitly, efforts are spend to bridge the distributions of L and H, provided that the computational costs of L and H are

sufficiently different, including additional steps that require sampling with L or a comparable level of theory does not significantly reduce the computational advantage of the multi-level free energy approach.

Given the fundamental aim of our approach, the key is then to identify the degrees of freedom responsible for creating the gap in the distributions of L and H. This is certainly not a trivial problem and can benefit from innovative approaches such as machine learning (see discussion in Sec. IV). For the case of integrating low-level and high-level QM potentials, which is the major motivation for our work, the likely situation is that they differ mainly in bonded degrees of freedom,[19,45,55] such as bond lengths, bond angles, and selected dihedrals. While low-level QM methods generally suffer from inadequate electronic polarization[59–61] and inaccurate multiple moments,[62,63] it is unlikely that the charge distribution is grossly misrepresented as compared to a high-level QM description; nevertheless, the impact of different non-bonded terms on the distribution is also explored in the following test cases. Denoting the degrees of freedom that cause a poor overlap between L and H as **X** and the rest (QM) degrees of freedom as **Y**, our strategy is then to treat them differently when converting between the L and H levels of theory.

B. A staged transformation approach

Recognizing that different "solute" degrees of freedom **X**, **Y** feature different distribution overlaps between L and H and that **X** tends to be more local in nature, the basic thermodynamic path that we propose is shown in Fig. 3(a) in which the system is transformed from L to H in a staged fashion: first, imagine turning on a confinement potential to restrain/constrain **X** to the free energy minimum on the L level ($\mathbf{X}_0$), then evaluate the "reorganization free energy cost" (ΔG_Λ^L) of changing $\mathbf{X}_0$ to $\mathbf{X}'_0$, the free energy minimum at the H level of theory, with **X** restrained/constrained to $\mathbf{X}'_0$, evaluate the free energy change associated with changing the potential function from L to H, $\Delta G_0^{L \to H}$, and finally, imagine turning off the confinement potential on **X** on the H level. Formally, the total free energy change is written as the sum of contributions from all stages (Fig. 3),
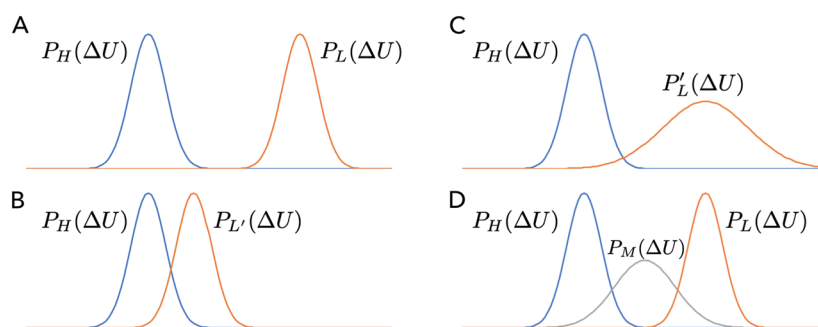
FIG. 2. Different strategies for improving the energy difference (ΔU) distribution overlap between L and H levels of theory, which is critical to the convergence of free energy simulations that rely on sampling at either one end-state or both end-states.[32,33,36,40] (a) The two end-states have a gap in the ΔU distribution. (b) One can modify the low-level method (L $\to$ L′) to improve the ΔU distribution based on, for example, force-matching[45,55] or a neural network.[51] (c) One can modify the distribution of the same low level method ($P_L \to P'_L$), with, for example, Tsallis sampling,[58] to improve the ΔU distribution. (d) One can introduce an intermediate level of method (or distribution) P_M to bridge the distributions at the L and H levels. The staged protocol described here falls into this category.
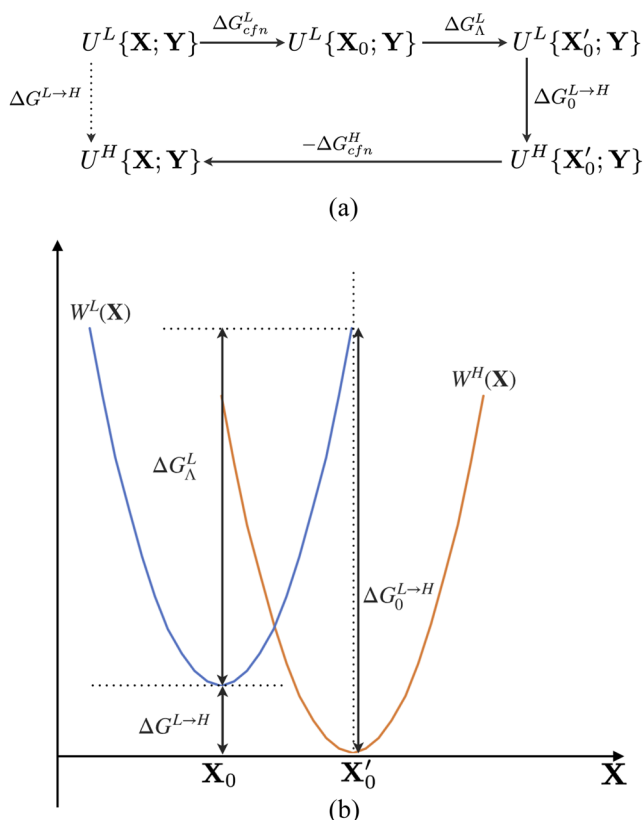
FIG. 3. A simple staged transformation approach for computing the free energy difference at two levels (L/H) of theory, $\Delta G^{L \to H}$. (a) The staged thermodynamic path treats the selected degrees of freedom ($\mathbf{X}$) separately from the rest ($\mathbf{Y}$); $\mathbf{X}$ represents the degrees of freedom that lead to a large gap in the ΔU^{LH} distribution. Assuming that the free energy costs for confining $\mathbf{X}$ to values at (or near) the free energy minima are similar at the L and H levels, $\Delta G^{L \to H}$, is given by the sum of $\Delta G_0^{L \to H}$, which converges readily since the sampling involves only $\mathbf{Y}$, and the "reorganization free energy," ΔG_Λ^L, which is the free energy cost of changing $\mathbf{X}_0$ to $\mathbf{X}_0'$ at the low level of theory. (b) A simple one-dimensional illustration for the staged transformation approach: $W(\mathbf{X})$ is the potential of mean force and the key approximation is that the configurational entropies for $\mathbf{X}$ are comparable at the L and H levels.

$$\Delta G^{L \to H} = \Delta G_{cfn}^L + \Delta G_\Lambda^L + \Delta G_0^{L \to H} - \Delta G_{cfn}^H. \qquad (1)$$

If we *approximate* that the confinement free energies evaluated at the L and H levels are comparable (*vide infra*), then the expression can be simplified into the sum of only two contributions,

$$\Delta G^{L \to H} \approx \Delta G_\Lambda^L + \Delta G_0^{L \to H}. \qquad (2)$$

The staging strategy can be better illustrated with a simple example where $\mathbf{X}$ is one-dimensional, as shown in Fig. 3(b). Define the potential of mean force in $\mathbf{X}$ at different levels (L/H) as

$$W^{L/H}(\mathbf{X}) = -k_B T \ln \int d\mathbf{Y} e^{-\beta U^{L/H}(\mathbf{X}, \mathbf{Y})} + C, \qquad (3)$$

we can express the free energy difference between L and H in terms of W at the free energy minimum and the entropic contributions associated with $\mathbf{X}$ ($S_\mathbf{X}^{L/H}$),

$$\Delta G^{L \to H} = [W^H(\mathbf{X}_0') - TS_\mathbf{X}^H] - [W^L(\mathbf{X}_0) - TS_\mathbf{X}^L], \qquad (4)$$

and here, the entropic contributions can be regarded as the confinement free energies in Eq. (1). If the entropic contributions at the L and H levels are comparable, this can be simplified and rearranged as

$$\begin{aligned} \Delta G^{L \to H} &\approx W^H(\mathbf{X}_0') - W^L(\mathbf{X}_0) \\ &= [W^H(\mathbf{X}_0') - W^L(\mathbf{X}_0')] - [W^L(\mathbf{X}_0) - W^L(\mathbf{X}_0')] \\ &\equiv \Delta G_0^{L \to H} + \Delta G_\Lambda^L. \end{aligned} \qquad (5)$$
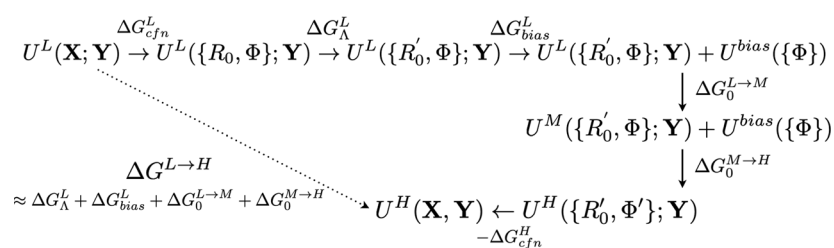
The significance of this simple rearrangement is that $\Delta G_0^{L \to H}$ is evaluated by sampling along the $\mathbf{Y}$ degrees of freedom, which feature a favorable overlap between L and H; the "reorganization" in the $\mathbf{X}$ degrees of freedom, ΔG_Λ^L, is evaluated at the low-level of theory.

Before discussing the evaluation of $\Delta G_0^{L \to H}$ and ΔG_Λ^L in detail in Subsection II C, we note that in these basic schemes discussed so far, we *approximate* that the confinement free energies ($\Delta G_{conf}^{L/H}$), which are mostly entropic in nature, are similar at the L and H levels of theory and therefore do not need to be evaluated explicitly. If the "solute" is structurally rigid, it is possible to explicitly evaluate the confinement free energies with normal mode analysis.[8,64,65] As the region of interest becomes large and/or flexible, either ignoring the difference between $\Delta G_{conf}^{L/H}$ and harmonic approximations becomes unreliable. Instead, it is preferable to limit $\mathbf{X}$ to a small number of bonded degrees of freedom for which free energy contributions can be evaluated using approximate analytic expressions[66] while treating differences in specific dihedral angles and/or non-bonded terms differently. Accordingly, we imagine a more general staged transformation path as illustrated in Fig. 4.

Along this more general transformation path, the difference in specific dihedral terms can be considered by turning on restraining potentials at the L level so that the corresponding dihedral distribution overlaps favorably with that at the H level; the free energy contribution of turning on the restraining potential can be evaluated explicitly as commonly done in binding free energy simulations.[67] In the event that the distribution overlap between L and H is also impacted by difference in non-bonded interactions, such as partial charge distributions, it is also possible to introduce an intermediate level of model M to bridge the distributions (see Sec. III).

C. Evaluation of $\Delta G_0^{L \to H}$ and ΔG_Λ^L

In this subsection, we discuss in more detail the calculations of $\Delta G_0^{L \to H}$ and ΔG_Λ^L. The former is done with $\mathbf{X}$ constrained at $\mathbf{X}_0'$, while the latter involves changing from $\mathbf{X}_0$ to $\mathbf{X}_0'$; thus, both calculations rely on the determination of free energy minima on the L and H free energy surfaces. Provided that $\mathbf{X}$ includes only localized degrees of freedom such as bond-stretches and angle-bending or stiff torsions, $\mathbf{X}_0$ and $\mathbf{X}_0'$ are expected to be well-defined and

$$U^L(\mathbf{X}; \mathbf{Y}) \xrightarrow{\Delta G^L_{cfn}} U^L(\{R_0, \Phi\}; \mathbf{Y}) \xrightarrow{\Delta G^L_\Lambda} U^L(\{R'_0, \Phi\}; \mathbf{Y}) \xrightarrow{\Delta G^L_{bias}} U^L(\{R'_0, \Phi\}; \mathbf{Y}) + U^{bias}(\{\Phi\})$$

$$\downarrow \Delta G_0^{L \to M}$$

$$U^M(\{R'_0, \Phi\}; \mathbf{Y}) + U^{bias}(\{\Phi\})$$

$$\downarrow \Delta G_0^{M \to H}$$

$$\begin{array}{c} \Delta G^{L \to H} \\ \approx \Delta G^L_\Lambda + \Delta G^L_{bias} + \Delta G_0^{L \to M} + \Delta G_0^{M \to H} \end{array} \qquad U^H(\mathbf{X}, \mathbf{Y}) \underset{-\Delta G^H_{cfn}}{\longleftarrow} U^H(\{R'_0, \Phi'\}; \mathbf{Y})$$

FIG. 4. A more general staging strategy for transforming from the L to H level of theory with the controlled distribution overlap along the path. Here, $\mathbf{X} = \{R, \Phi\}$; the general aim is to minimize the number of degrees of freedom that are constrained ($\{R\}$) such that the approximation of similar "confinement free energies" at the L/H levels is expected to hold; any difference in softer degrees of freedom, such as selected dihedrals ($\{\Phi\}$), can be controlled with a biasing potential and the corresponding free energy can be evaluated as commonly done in ligand-binding simulations.[67] To further improve the distribution overlap, an intermediate level of model with improved partial charges, for instance, can be introduced (see Sec. III).

straightforward to determine with relatively short equilibrium MD simulations (see Sec. III B for a discussion of impacts of choosing different $\mathbf{X}_0$ and $\mathbf{X}'_0$ values and Sec. III C for a discussion of sampling requirement).

Once $\mathbf{X}_0$ and $\mathbf{X}'_0$ are determined, $\Delta G_0^{L \to H}$ and ΔG^L_Λ can be computed using the well-established simulation methodologies. Specifically, for $\Delta G_0^{L \to H}$, since sampling along $\mathbf{Y}$ is expected to feature a favorable distribution overlap between L and H, it is possible that 1-step free energy perturbation that relies on sampling only at the L level is sufficient for practical convergence. On the other hand, it is always more robust to conduct some sampling at the H level so that the distribution overlap can be explicitly assessed, including sampling at the H level also enables more reliable free energy estimates using techniques such as the linear response approximation (LRA)[18,20,68] or Bennet acceptance ratio (BAR).[69–71]

For the computation of ΔG^L_Λ, which involves sampling only at the L level, different strategies can be employed. The most straightforward approach is to use the dual topology free energy perturbation to convert the solute/QM region from $\mathbf{X}_0$ to $\mathbf{X}'_0$. One could pursue an alternative thermodynamic path (see Fig. 6) in which the solute, while adopting structure $\mathbf{X}_0$ and $\mathbf{X}'_0$, is first decoupled from the environment; then, the conversion from $\mathbf{X}_0$ to $\mathbf{X}'_0$ can be done with free energy perturbation in the gas phase, which is expected to converge rapidly. The important point is that regardless of the chosen path, all ΔG^L_Λ simulations require sampling at the L level only and therefore are computationally inexpensive.

D. Computational setup of test systems

As a proof of concept, we illustrate the staged transformation approach with two simple but non-trivial examples in a water droplet (Fig. 5) and we examine a single vertical leg of the thermodynamic cycle in Fig. 1. For the purpose of this work, we mainly focus on the choice of CHARMM36[72] as the low (L) level and DFTB3/3OB/MM[73] as the high (H) level; we also study the case in which L is DFTB3/3OB/MM and H is B3LYP/6-31G(d)/MM.

1. System setup

The small molecule solute is weakly restrained to be at the center of a 12 Å radius water droplet, which is not subject to any special solvent boundary potential. Non-bonded interactions at the pure MM level are computed with extended electrostatics[74] available in CHARMM,[75] which computes electrostatic interactions with group-based multipoles beyond the cutoff (12 Å) distance; this option is consistent with the fact that QM/MM electrostatic interactions are computed without any cutoff.

For simulations aimed at probing the distribution overlap between L = MM (CHARMM36) and H = DFTB3/3OB/MM, the molecular dynamics simulations are typically NVT simulations at 300 K with ~100 ps of sampling. For L = DFTB3/3OB/MM and H = B3LYP/MM, the sampling is ~50 ps. For the integration time step, it is 0.5 fs for simulations with a completely unrestrained solute. With the bonded degrees of freedom constrained (SHAKE[76]), the integration time step is 2.0 fs. For simulations that also probe the effect of restraining angular degrees of freedom, the force constant used is 300 kcal/mol rad^2.

2. Free energy simulations

With L = MM (CHARMM36) and H = DFTB3/3OB/MM, it is possible to compute the $\Delta G^{L \to H}$ with "brute-force" thermodynamic integration using multiple λ windows that connect the L and H levels. We have done so for the cases of unrestrained solutes and for each solute, 11 λ windows (0.0, 0.05, 0.15, . . ., 0.95, 1.0) are used with each window sampled by 350 ps of production runs followed by 60 ps of equilibration. For the more approximate free energy estimators: 1-λ FEP, LRA, and BAR, 100 ps of production run is used for the relevant windows; 1-λ FEP employs only the L-level sampling, while the other estimators employ sampling with both end-states (L and H). The lengths of these simulations are chosen such that they represent the typically affordable level of computations using QM potentials such as DFT. In addition to unrestrained solutes, we have also examined the consistency among these approximate free energy estimators for cases where the solute is either completely frozen or partially constrained/restrained (see Tables for details); the sampling is also 100 ps of production run for the relevant windows. Finally, for L = DFTB3/3OB/MM and H = B3LYP/MM, we also computed $\Delta G^{L \to H}$ using the various approximate free energy estimators with 50 ps of production for the relevant windows.

For the exploration of the staged transformation protocol, we focus on the case of unrestrained solutes with L = MM
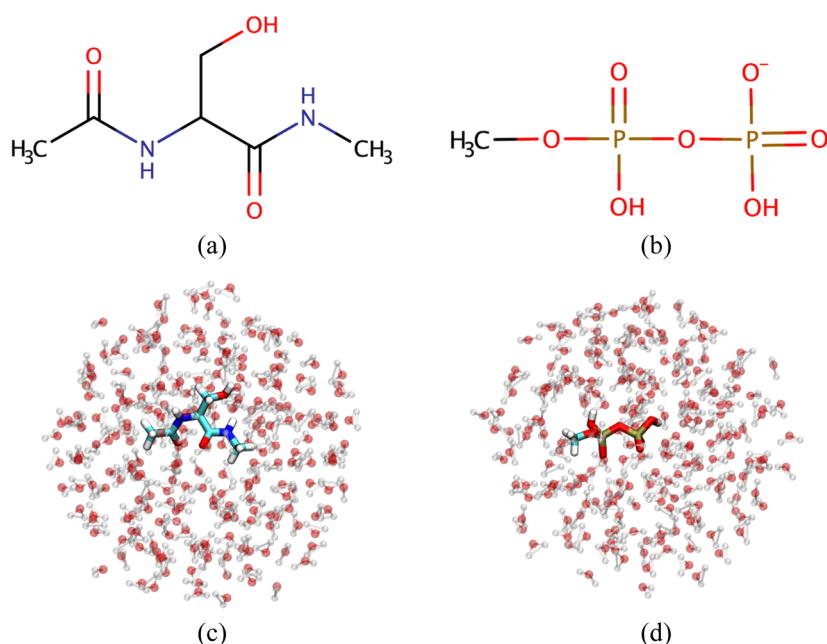
FIG. 5. Test systems explored in this work: (a) blocked serine: *N*-acetyl-serine-methyl-amide (Ser) and (b) methyl di-phosphate (MDP). [(c) and (d)] A water droplet of radius 12 Å that contains a Ser or MDP.

(CHARMM36) and H = DFTB3/3OB/MM. For the "confinement step," we apply both bond constraint (SHAKE) and angle restraint using a large force constant of 300 kcal/mol rad^2 with the reference structures collected from free 1 ns L (i.e., $\mathbf{X}_0$) and H (i.e., $\mathbf{X}_0'$) simulations (see Sec. III B for a discussion of selection criteria); to test the sensitivity of the results to the choice of the reference, we choose five $\mathbf{X}_0$ and five $\mathbf{X}_0'$, leading to 25 possible combinations. As discussed below, even with the constraints/restraints on the bonded degrees of freedom, the overlap between the L and H distributions is not optimal. Therefore, we take the advantage of the flexibility of the staged transformation approach (Fig. 4) and include an additional intermediate (M) level, which differs only from L in terms of the partial charges; instead of the standard CHARMM36 charges, we use the Mulliken charges at the DFTB3/3OB level calculated for a random snapshot collected from the aforementioned DFTB3/3OB/MM simulations.

For the computation of ΔG_Λ^L, since L = CHARMM36, we use the specific thermodynamic cycle outlined in Fig. 6, which involves five separate sets of free energy simulations in either the water droplet [$\Delta G_{slv}^L(R_0)$ and $\Delta G_{slv}^L(R_0')$] or in the gas phase [$\Delta G_{nbond}^L(R_0)$, $\Delta G_{nbond}^L(R_0')$, and $\Delta G_{bonded}^L(R_0, R_0')$]. Since all these simulations are at the low-level, they are readily conducted with multi-λ thermodynamic integration with 11 λ windows (0.0, 0.05, 0.15, . . ., 0.95, 1.0), with each window sampled for at least 100 ps.

3. Scores used to evaluate distribution overlap

To help characterize the distribution overlap between the L and H levels, we follow the discussions of Kofke and co-workers.[37–40,77,78] In particular, we have computed the following two sets of scores. The first set is based on the total energy distributions at different levels of theories. As discussed in Refs. 38 and 39, we define the following energy distributions:

- $\rho_{LL}(U_L) \equiv \rho_L$: sample with L, examine the distribution of U_L;
- $\rho_{HH}(U_H) \equiv \rho_H$: sample with H, examine the distribution of U_H;
- $\rho_{LH}(U_H)$: sample with L, examine the distribution of U_H; and
- $\rho_{HL}(U_L)$: sample with H, examine the distribution of U_L.

Then, one can define the following overlap integrals, which range between 0 and 2 depending on the degree of overlap: (a) H in L,

$$K_{HL} = 2 \int_{-\infty}^{\infty} dU_L \rho_{LL}(U_L) \int_{-\infty}^{U_L} dU_L' \rho_{HL}(U_L'). \quad (6)$$
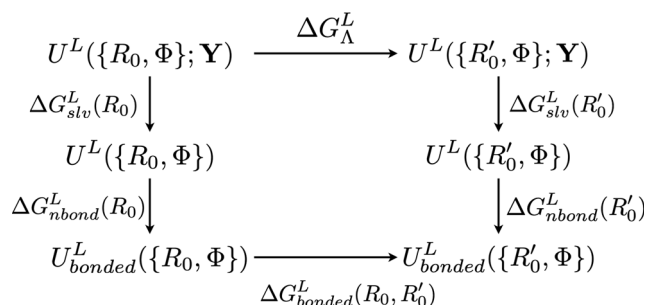
FIG. 6. A sample thermodynamic cycle used to compute the reorganization free energy, ΔG_Λ^L, in which the solute is first decoupled from the environment and then removed of any non-bonded interactions; finally, the difference in the bonded free energies is evaluated, leading to $\Delta G_\Lambda^L = [\Delta G_{slv}^L(R_0') - \Delta G_{slv}^L(R_0)] + [\Delta G_{nbond}^L(R_0') - \Delta G_{nbond}^L(R_0)] + \Delta G_{bonded}^L(R_0, R_0')$. Note that all calculations are done at the low level of theory, and most steps are done in the gas phase; thus, the computational cost is modest.

(b) L in H,

$$K_{LH} = 2 \int_{-\infty}^{\infty} dU_H \rho_{HH}(U_H) \int_{-\infty}^{U_H} dU'_H \rho_{LH}(U'_H). \quad (7)$$

The second set of scores is based on the Kullback–Leibler (KL) divergence of total energy distributions,

$$s_L = \int_{\Gamma} d\gamma \rho_L(\gamma) ln \frac{\rho_L(\gamma)}{\rho_H(\gamma)}, \quad (8)$$

$$s_H = \int_{\Gamma} d\gamma \rho_H(\gamma) ln \frac{\rho_H(\gamma)}{\rho_L(\gamma)}. \quad (9)$$

These can be shown to be related to the dissipative work associated with the L → H conversion,

$$s_L = \langle \beta W_{L \to H} \rangle_L - \beta \Delta F, \quad (10)$$

$$s_H = -\langle \beta W_{H \to L} \rangle_H + \beta \Delta F, \quad (11)$$

where W is the work; for FEP, it is just the energy gap ΔU_{HL}. These relative entropies can be used to define the Π scores,

$$\Pi_{LH} = \sqrt{\frac{s_L}{s_H} \mathbf{W}_L \left(\frac{1}{2\pi} (M-1)^2 \right)} - \sqrt{2s_L}, \quad (12)$$

$$\Pi_{HL} = \sqrt{\frac{s_H}{s_L} \mathbf{W}_L \left(\frac{1}{2\pi} (M-1)^2 \right)} - \sqrt{2s_H}, \quad (13)$$

where M is the sample size, $\mathbf{W}_L(x)$ is the Lambert W function, defined as the solution for w in the equation $x = we^w$. It was proposed that good sampling is expected if $\Pi > 0$. Woodcock and Boresch explored a simplified version of the Π score that assumes $s_L = s_H$.[41]

III. RESULTS AND DISCUSSION

In the following, we first illustrate the impact of separately treating **X** degrees of freedom on the distribution overlap between L and H, then we demonstrate that the staged transformation approach indeed leads to the expected free energy difference between L and H by comparing to brute-force thermodynamic integration for $\Delta G^{L \to H}$, which involves explicitly mixing the L and H potential functions in multiple-λ windows. Finally, we briefly comment on the issue of sampling on both the L and H levels of potential energy surfaces. While the work was motivated by conducting multi-level QM/MM simulations, the L level is taken to be CHARMM36 in most analyses, with DFTB3/3OB/MM being the H level; nevertheless, we also include results on the combination of DFTB3/3OB/MM as L and B3LYP/6-31G(d)/MM as H.

A. Improvement of distribution overlap with confinement of X degrees of freedom

As discussed extensively in the literature,[32,40] free energy simulation results depend critically on the ΔU distribution
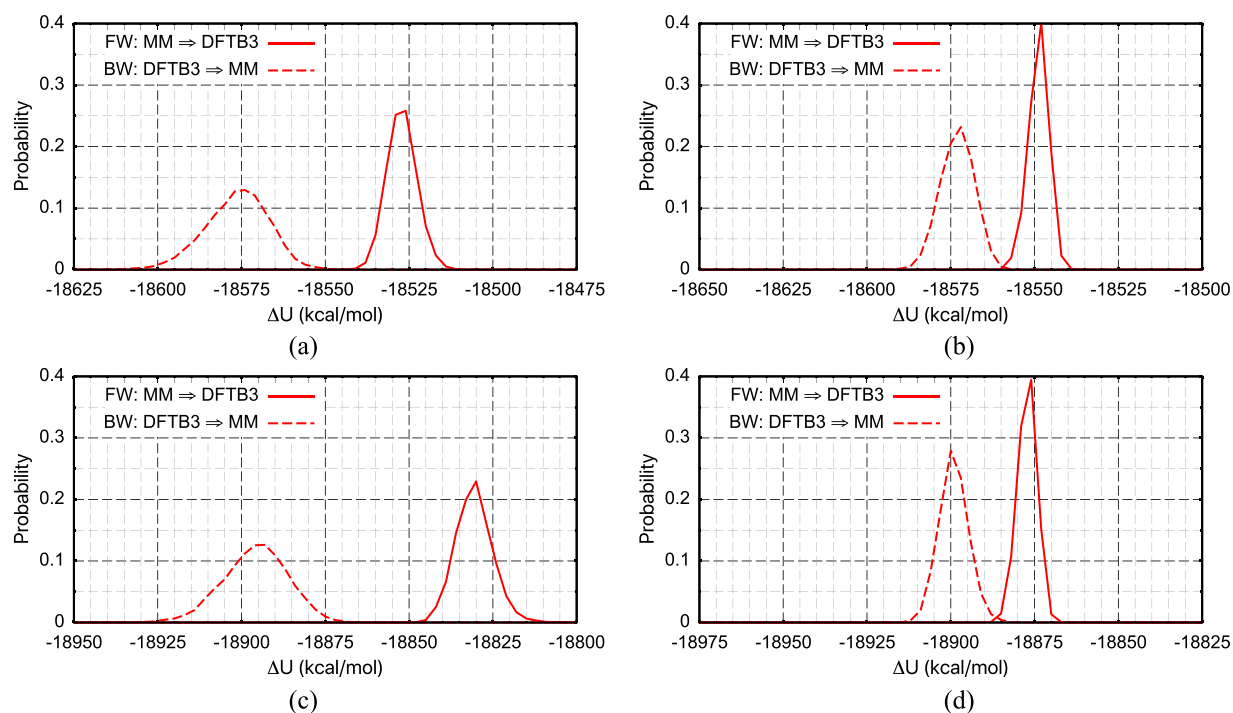
FIG. 7. Energy difference (ΔU^{LH}) distribution for L = CHARMM36 and H = DFTB3/3OB/MM. The solid lines are for the "forward (FW)" distribution (i.e., sampling with L), and the dashed lines are for the "backward (BW)" distribution (i.e., sampling with H). The four solution cases are (a) a fully unrestrained blocked serine, (b) a blocked serine with a frozen structure, (c) a fully unrestrained methyl diphosphate (MDP), and (d) a MDP with a frozen structure.

TABLE I. Computed $\Delta G^{L \to H}$ values (in kcal/mol) and distribution overlap scores for a blocked serine in water.[a]

			ΔG[b]					Overlap scores[c]				
Low	High	Constraints	M-λ TI	1-λ FEP	LRA	BAR	Staged	Π_{LH}	Π_{HL}	Overlap (%)	K_{HL}	K_{LH}
C36	DFTB3	Free	−18 545.3	−18 541.2 (0.6)	−18 545.5 (0.5)	. . .	−18 545.4 (2.9/1.2)	−4.04	−4.09	0.01	0.58	0.30
C36	DFTB3	Frozen	. . .	−18 557.1 (0.6)	−18 559.1 (0.5)	−18 559.1 . . .	. . .	−2.03	−1.64	0.22	0.79	0.55
MULL	DFTB3	Bond	. . .	−18 484.5 (0.3)	−18 485.4 (0.2)	−18 485.9 . . .	. . .	−0.89	0.34	2.01	0.95	0.70
MULL	DFTB3	Frozen	. . .	−18 483.6 (0.2)	−18 483.7 (0.2)	−18 483.6 (0.2)	. . .	1.66	1.42	31.67	0.92	0.82
DFTB3	B3LYP	Free	. . .	−339 808.6 (0.6)	−339 809.1 (0.5)	−339 809.0 . . .	. . .	−0.17	−0.68	2.75	0.73	0.43

[a]C36 = CHARMM36; DFTB3 = DFTB3/3OB/MM; and in MULL, the partial charges of the solute are taken as Mulliken charges from a DFTB3 calculation and held fixed during each MULL simulation.
[b]M-λ TI: 11-λ window thermodynamic integration; 1-λ FEP: single window FEP based on L sampling; LRA: linear response approximation based on sampling with both end-states (L and H) windows; BAR: Bennet-Acceptance-Ratio based on the same sampling data from LRA; and Staged: staged transformation protocol introduced here. The statistical errors are estimated based on block averaging and are less than 0.1 kcal/mol for M-λ TI. For the staged transformation, the uncertainty is estimated based on 25 different choices of $\mathbf{X}_0$, $\mathbf{X}_0'$ combinations (see Table III for detailed data).
[c]$\Pi_{LH,HL}$: Eqs. (12) and (13); $K_{HL,LH}$: Eqs. (6) and (7); and Overlap %: the fraction of ΔU data points that lie in the overlapped region.

overlap of the two end-states. The challenge of establishing the good distribution overlap is well illustrated in Figs. 7(a) and 7(c): without any conformational restraint, there is essentially no ΔU overlap between the CHARMM36 (L) and DFTB3/3OB/MM (H) simulations for both a serine and a methyl diphosphate (MDP) in solution. The overlap is even smaller for the case of MDP with an apparent gap in the ΔU distribution of ~25 kcal/mol. The lack of overlap is well captured by the various overlap scores reported in Tables I and II, for example, the Π scores are large and negative with values ~−4.

To reveal whether the lack of overlap is largely dominated by the intra-solute degrees of freedom or by the non-bonded parameters at the low (CHARMM36) level, we first reexamine the ΔU distributions with simulations in which the solute structure is frozen. As shown in Figs. 7(b) and 7(d), freezing the intra-solute degrees of freedom improves the ΔU overlap for both cases analyzed; the improvement is visibly more transparent for MDP, suggesting that the CHARMM36-CGENFF[79] model leads to considerably different conformations for MDP compared to DFTB3/3OB. Examination of snapshots suggests that the difference largely lies

TABLE II. Computed $\Delta G^{L \to H}$ values (in kcal/mol) and distribution overlap scores for a methyl diphosphate in water.[a]

			ΔG					Overlap scores				
Low	High	Constraints	M-λ TI	1-λ FEP	LRA	BAR	Staged	Π_{LH}	Π_{HL}	Overlap %	K_{HL}	K_{LH}
C36	DFTB3	Free	−18 857.3	−18 845.1 (0.5)	−18 858.2 (0.6)	. . .	−18 858.0 (1.9)	−3.92	−3.92	0.00	0.22	0.45
C36	DFTB3	Frozen	. . .	−18 885.0 (0.7)	−18 885.9 (0.6)	−18 885.9 . . .	. . .	−2.00	−1.52	0.70	0.75	0.68
MULL	DFTB3	B + A	. . .	−18 813.3 (0.3)	−18 814.9 (0.3)	−18 814.9 . . .	. . .	−0.22	0.88	2.11	0.77	0.93
MULL	DFTB3	Frozen	. . .	−18 820.4 (0.1)	−18 820.3 (0.1)	−18 820.3 . . .	. . .	2.69	2.95	84.75	0.81	0.83
DFTB3	B3LYP	Free	. . .	−765 491.4 (0.3)	−765 494.1 (0.3)	−765 493.8 . . .	. . .	−1.20	−1.63	0.17	0.33	0.02
DFTB3	B3LYP	B + A	. . .	−765 501.7 (0.2)	−765 502.6 (0.2)	−765 502.3 (0.2)	. . .	0.66	0.09	5.36	0.80	0.90

[a]See footnote of Table I for notations; B+A means constrained bond lengths and restrained bond angles with a force constant of 300 kcal/mol rad[2]. For detailed data for the staged transformation with different $\mathbf{X}_0$, $\mathbf{X}_0'$ combinations, see Table IV.

in the bond angles and several dihedrals. Correspondingly, various overlap scores improve significantly; the Π scores, for example, increase from ∼−4 for the unconstrained simulations to ∼−2 for the frozen-solute simulations.

Even with the solute structure entirely frozen (thus fully consistent between L and H simulations), the ΔU distribution overlap is minimal, suggesting that there are significant differences in the non-bonded interactions between the solute and the solvent in the L and H simulations. The situation can be substantially improved by replacing the MM partial charges at the L level with Mulliken charges calculated at the DFTB3/3OB/MM level based on *a single random snapshot*, as shown in Figs. 8(a) and 8(b) for serine and MDP, respectively; the Π scores are now positive and in the range of ∼1.5 to 3. Since the Mulliken charges in these intermediate models are *not* updated during the simulation, the excellent ΔU overlaps suggest that the lack of electronic polarization *per se* is not a major factor that impacts the distribution overlap, even for a highly charged and polarizable solute such as MDP. However, the overlap deteriorates significantly when the solute structure is no longer frozen with only bonds constrained with SHAKE [Figs. 8(c) and 8(d)], highlighting the importance of capturing the structural dependence of charge distribution. In the case of MDP, constraining the covalent bonds and restraining the bond angles help improve the ΔU distribution overlap considerably.

When the L is DFTB3/3OB/MM and H is B3LYP/6-31G(d)/MM, there is a decent (∼3%) overlap in ΔU distribution for serine in water, but the overlap is minimal (∼0.17%) for MDP [Figs. 9(a) and 9(b)]; the Π scores are small negative values and ∼−1, respectively. With constrained bond length and restrained bond angles, as shown in Fig. 9(c), the overlap improves substantially for MDP to ∼5.4%; the Π scores also become positive (Table II).

In short, the results in this subsection indicate that both bonded and non-bonded degrees of freedom may contribute to modulate the distribution overlap between L and H levels of theory. With a decent QM level as L, it is expected that the overall charge distribution is reasonably represented compared to H, even for a highly charged and polarizable molecule such as MDP. Small differences in stiff bonded degrees of freedom such as bond stretch and bond angle can lead to a poor distribution overlap, which can be improved with constraints/restraints. When the L level is MM, the non-bonded interactions might be sufficiently different from the H level, and thus, constraining/restraining stiff intra-solute degrees of freedom alone may not be sufficient for improving the distribution overlap to a satisfactory level. For the test cases studied here, replacing the MM charges with a single set of Mulliken charges improves the overlap significantly, especially when combined with constraining/restraining stiff intra-solute degrees of freedom.
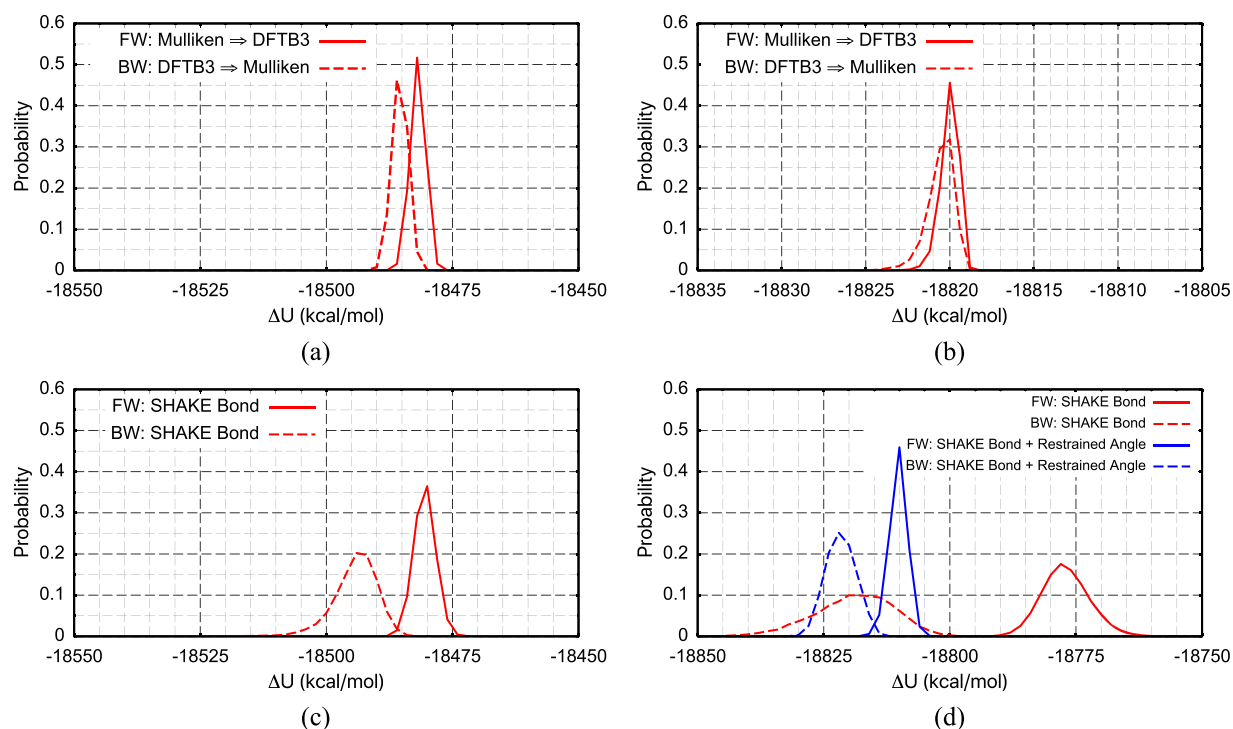
FIG. 8. Similar to Fig. 7, but the partial charges on the MM solute are replaced by Mulliken charges from a DFTB3/3OB/MM calculation; note that these Mulliken charges are held fixed in each trajectory. The four solution cases are (a) a blocked serine with a frozen structure, (b) a MDP with a frozen structure, (c) a blocked serine with all covalent bonds constrained with SHAKE, and (d) a MDP with all covalent bonds constrained using SHAKE and free (red) or restrained (blue) bond angles.
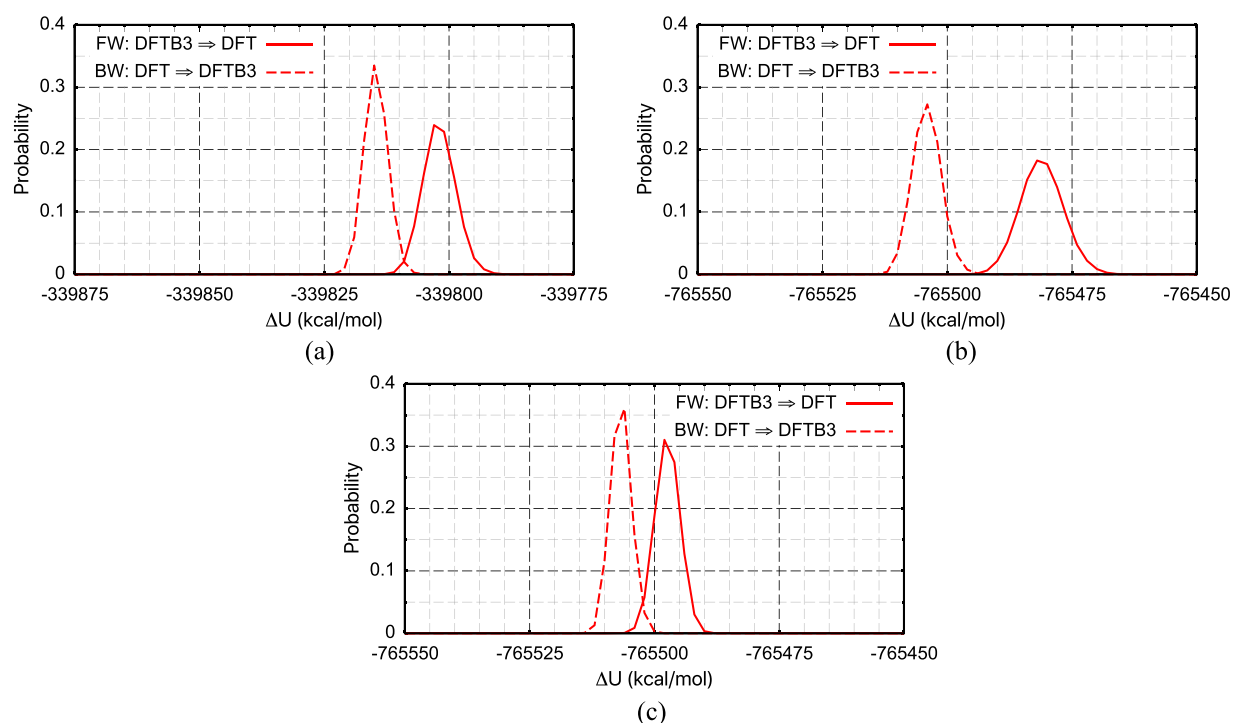
FIG. 9. Similar to Fig. 7 but with L = DFTB3/3OB/MM and H = B3LYP/6-31G(d)/MM. The three solution cases studied are (a) a fully unrestrained blocked serine, (b) a fully unrestrained MDP, and (c) a MDP with covalent bonds constrained using SHAKE and bond angles restrained.

## B. Staged transformation is able to complete the thermodynamic cycle

In Sec. III A, we examine whether the staged transformation approach is able to reproduce the $\Delta G^{L \to H}$ result of brute-force multi-λ thermodynamic integration for free serine and MDP in solution. Prior to the discussion, we first compare $\Delta G^{L \to H}$ computed with various approximations for the different simulations used to explore the ΔU overlap in the last subsection; this helps better illustrate the impact of the distribution overlap on the reliability of estimated $\Delta G^{L \to H}$.

With a free solute in water, the distribution overlap is poor for serine and MDP for both L/H combinations tested here; accordingly, the simple 1-λ FEP result is significantly different from other estimators (Tables I and II). For a free serine in water, L = CHARMM36 and H = DFTB3/3OB/MM, the 1-λ FEP value differs by ~4 kcal/mol from the multi-λ TI result; the difference is even larger, ~13 kcal/mol for MDP. With L = DFTB3/3OB/MM and H = B3LYP/6-31G(d)/MM, the overlap is minimal for MDP in water; thus, 1-λ FEP differs from a linear response approximation (LRA) estimator by ~3 kcal/mol. By contrast, the overlap is appreciable (~2.75%) for a free serine in water, and 1-λ FEP agrees with LRA to ~0.5 kcal/mol.

For all other cases explored here, which involves either changing the MM charges or confining/restraining stiff intra-solute degrees of freedom, improvement of the ΔU overlap is highly correlated with the level of consistency among 1-λ FEP, LRA, and BAR estimators. As far as the percentile of the overlap is higher than 1%, the three estimators are within 1 kcal/mol (see Tables I and II). Another observation is that the LRA estimator is robust for the two test systems studied here. Even for the cases of free solutes in water for which the distribution overlap between L/H is poor, the LRA estimator is ~1 kcal/mol from the brute-force multi-λ simulation; this is rather remarkable considering that the error for 1-λ FEP can be as large as 13 kcal/mol. This observation resonates with several previous studies that also highlighted the significant gain in conducting sampling on the H level,[18,20,68] even in cases where BAR has difficulty in convergence.

To examine the effectiveness and limitations of the staged transformation approach, we study the cases of fully unrestrained small solutes (serine and MDP) in water with L = CHARMM36 and H = DFTB3/3OB/MM for which the reference $\Delta G^{L \to H}$ values are readily computed using brute-force multi-λ TI simulations. From the results shown in Tables I and II, the staged transformation approach indeed gives an average result consistent with the multi-λ TI reference value for both Ser and MDP in water; the difference appears to be less than 1 kcal/mol. However, we note that the staged transformation approach requires the determination of equilibrium or representative structures at the L ($\mathbf{X}_0$) and H ($\mathbf{X}_0'$) levels in the "confinement step." Therefore, it is important to understand the sensitivity of the result to the choice of these reference structures,
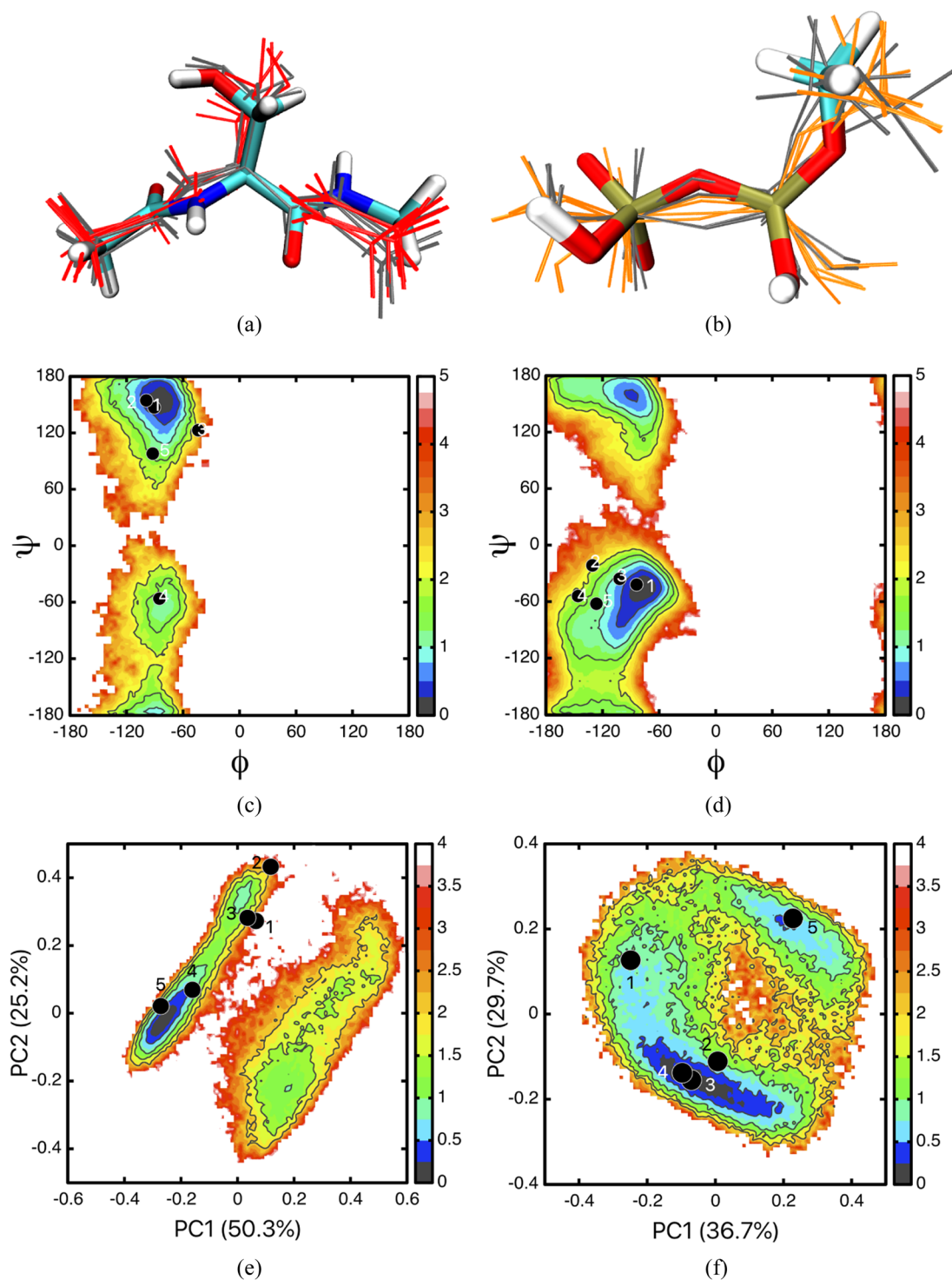
FIG. 10. Distribution of $\mathbf{X}_0$, $\mathbf{X}_0'$ for a blocked serine and MDP. (a) and(b) illustrate the superposition of the structures; one $\mathbf{X}_0$ is shown in licorice and colored based on element, the rest $\mathbf{X}_0$ shown in gray lines; $\mathbf{X}_0'$ structures are shown in a different color in line form. [(c) and (d)] $\Phi - \Psi$ free energy surfaces of a blocked serine in solution with (c) CHARMM36 and (d) DFTB3-Mulliken charges, the latter better resembles the DFTB3/3OB/CHARMM result (see, e.g., Ref. 53). [(e) and (f)] Free energy map projected onto the first two principal components for a MDP in solution using (e) CHARMM36 and (f) DFTB3/3OB/CHARMM. The dots indicate the locations of $\mathbf{X}_0$, $\mathbf{X}_0'$, and the labels correspond to those used in Table III.

TABLE III. Computed $\Delta G^{L \to H}$ values (in kcal/mol) for a fully unrestrained blocked serine in solution from the staged transformation approach using different combinations of confined reference structures (i.e., $\mathbf{X}_0$, $\mathbf{X}_0'$).[a]

	$\mathbf{X}_0'$ sample 1	$\mathbf{X}_0'$ sample 2	$\mathbf{X}_0'$ sample 3	$\mathbf{X}_0'$ sample 4	$\mathbf{X}_0'$ sample 5
$\Delta G_0^{L \to H}$	$-18\,550.6 \pm 0.3$	$-18\,551.9 \pm 0.2$	$-18\,553.9 \pm 0.3$	$-18\,556.3 \pm 0.5$	$-18\,550.2 \pm 0.2$
$\mathbf{X}_0$ sample 1	4.5/**$-18\,546.1$**[b]	5.0/$-18\,546.8$	9.6/$-18\,544.3$	11.1/$-18\,545.2$	11.1/$-18\,539.0$
$\mathbf{X}_0$ sample 2	2.4/$-18\,548.2$	2.9/$-18\,548.9$	7.6/$-18\,546.3$	9.0/$-18\,547.3$	9.1/$-18\,541.1$
$\mathbf{X}_0$ sample 3	3.9/$-18\,546.7$	4.5/$-18\,547.4$	9.1/$-18\,544.8$	10.6/$-185\,45.7$	10.6/$-18\,539.6$
$\mathbf{X}_0$ sample 4	3.1/$-18\,547.5$	3.7/$-18\,548.2$	8.3/$-18\,545.7$	9.7/$-18\,546.5$	9.8/$-18\,540.4$
$\mathbf{X}_0$ sample 5	3.0/$-18\,547.6$	3.5/$-18\,548.3$	8.2/$-18\,545.8$	9.6/$-18\,546.7$	9.6/$-18\,540.5$
Avg.[c]	$-18\,545.4 \pm 2.9$	$(-18\,546.7 \pm 1.2)$			

[a]See text for the discussion of the selection of the confined structures. The first row lists the values of $\Delta G_0^{L \to H}$, which depends only on the choices of R_0', as discussed in Sec. II D 2, we introduced an intermediate level of model (M) that uses the CHARMM36 force field, except that the partial charges are replaced by DFTB3/3OB Mulliken charges computed for a single snapshot in solution; thus, $\Delta G_0^{L \to H} = \Delta G_0^{L \to M} + \Delta G_0^{M \to H}$, where the L/M conversion is done with multi-λ TI due to the low computational cost, and M/H conversion is done with 1-λ FEP due to the favorable distribution overlap illustrated in Fig. 8. For other entries, the value before the slash is ΔG_Λ^L, and the value after the slash is $\Delta G^{L \to H} = \Delta G_0^{L \to H} + \Delta G_\Lambda^L$.
[b]For this specific combination, the confined structures ($\mathbf{X}_0$, $\mathbf{X}_0'$) are close to the minima on the L/H free energy surface based on the 2D projection [see Figs. 10(c) and 10(d)].
[c]The average $\Delta G^{L \to H}$ value is reported as the "Stage" entry in Table I; values with parentheses are the results without including data from R_0' sample 5 [see Fig. 10(d)], which leads to consistently different values (see text).

especially if the confinement free energies $\Delta G_{cfn}^{L/H}$ are not explicitly computed, i.e., the approximation is that they cancel out perfectly between the L and H levels.

To this end, for each test case, we have sampled 25 different combinations of $\mathbf{X}_0$ and $\mathbf{X}_0'$ by choosing five reference structures at each level of theory. Ideally, $\mathbf{X}_0$ and $\mathbf{X}_0'$ should be selected by identifying structures near the free energy minima at L and H levels of theory [Fig. 3(b)], which requires projecting the free energy surface onto different dimensions using, for example, principal component analysis.[80] In this work, to test the robustness of the result, we chose $\mathbf{X}_0$ and $\mathbf{X}_0'$ by simply identifying structures that have the lowest potential energies for the solute, including both intra-solute and solute–solvent interactions. As illustrated in Figs. 10(c)–10(f), some of the selected structures are indeed close to the free energy minima, while others are not. The spread of the conformations, as shown by the superpositions in Figs. 10(a) and 10(b), is more modest for the blocked serine than MDP; the root mean squared difference (RMSD) among the different structures is ~0.05 Å, with some dihedral angles differing by ~20° for the blocked serine and ~40° for MDP. In

principle, the staged transformation approach ought to be applicable to any reference structure ($\mathbf{X}_0$ and $\mathbf{X}_0'$) combinations, although the assumption of canceled $\Delta G_{cfn}^{L/H}$ is expected to work better for $\mathbf{X}_0$ and $\mathbf{X}_0'$ close to the respective free energy minimum.

The results for these 25 combinations are shown in Tables III and IV, which illustrate several points worth noting. First, without considering the free energy components associated with changing the confined degrees of freedom from $\mathbf{X}_0$ to $\mathbf{X}_0'$, the free energy quantity $\Delta G_0^{L \to H}$ is sensitive to the choice of $\mathbf{X}_0'$; the magnitude of variation is ~6 kcal/mol for serine and ~8 kcal/mol for MDP. More importantly, the magnitude of $\Delta G_0^{L \to H}$ can be substantially different from the true $\Delta G^{L \to H}$ value, by as much as ~20 kcal/mol for MDP. Indeed, the "reorganization free energy," ΔG_Λ^L, ranges from 3 kcal/mol to 11 kcal/mol for the blocked serine, and 12 kcal/mol to 25 kcal/mol for MDP (see numbers before slashes in Tables III and IV). Second, for the combination of $\mathbf{X}_0$, $\mathbf{X}_0'$ that are close to the free energy minima (highlighted in bold in Tables III and IV), at least based on the 2D projection in Fig. 10, the computed $\Delta G^{L \to H}$ values are indeed close to the m-λ TI results ($-18\,545.3$ kcal/mol

TABLE IV. Computed $\Delta G^{L \to H}$ values (in kcal/mol) for a fully unrestrained methyl diphosphate in solution from the staged transformation approach using different combinations of confined reference structures (i.e., $\mathbf{X}_0$, $\mathbf{X}_0'$).[a]

	$\mathbf{X}_0'$ sample 1	$\mathbf{X}_0'$ sample 2	$\mathbf{X}_0'$ sample 3	$\mathbf{X}_0'$ sample 4	$\mathbf{X}_0'$ sample 5
$\Delta G_0^{L \to H}$	$-18\,878.6 \pm 0.3$	$-18\,878.6 \pm 0.2$	$-18\,876.4 \pm 0.2$	$-18\,872.8 \pm 0.2$	$-18\,880.5 \pm 0.3$
$\mathbf{X}_0$ sample 1	23.7/$-18\,854.8$	21.6/$-18\,857.0$	17.7/$-18\,858.7$	16.0/$-18\,856.8$	24.7/$-18\,855.8$
$\mathbf{X}_0$ sample 2	22.9/$-18\,855.7$	20.8/$-18\,857.8$	16.9/$-18\,859.5$	15.2/$-18\,857.6$	23.9/$-18\,856.6$
$\mathbf{X}_0$ sample 3	20.0/$-18\,858.6$	17.8/$-18\,860.7$	13.9/$-18\,862.4$	12.3/$-18\,860.5$	20.9/$-18\,859.6$
$\mathbf{X}_0$ sample 4	23.2/$-18\,855.3$	21.1/$-18\,857.5$	17.2/$-18\,859.1$	15.5/$-18\,857.2$	24.2/$-18\,856.3$
$\mathbf{X}_0$ sample 5	21.7/$-18\,856.9$	19.6/$-18\,859.0$	15.7/**$-18\,860.7$**[b]	14.0/**$-18\,858.8$**[b]	22.6/$-18\,857.8$
Avg.	$-18\,858.0 \pm 1.9$				

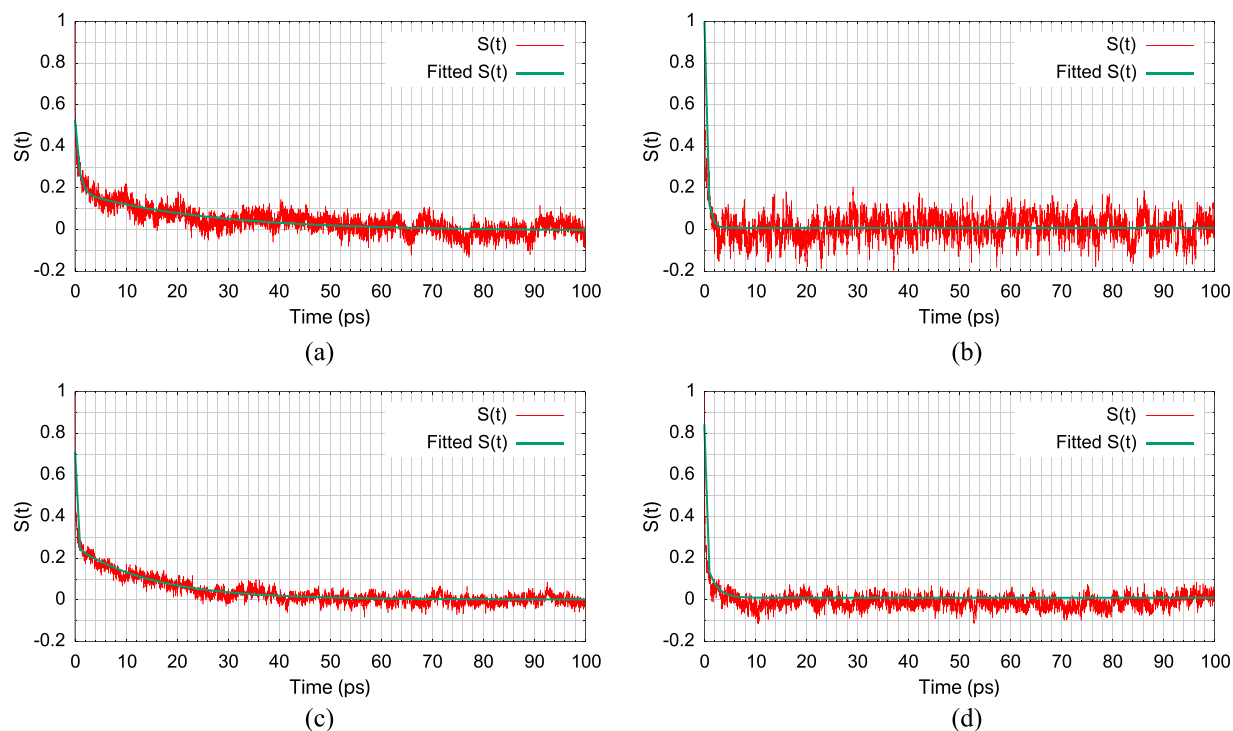[a]See footnote a of Table III for the format of the entries. The average $\Delta G^{L \to H}$ value is reported as the "Stage" entry in Table II.
[b]For these specific combinations, the confined structures ($\mathbf{X}_0$, $\mathbf{X}_0'$) are close to the minima on the L/H free energy surface based on the 2D projection [see Figs. 10(e) and 10(f)].
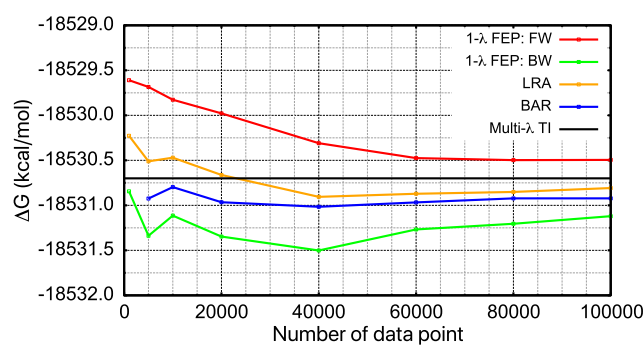
for serine and $-18\,857.3$ kcal/mol for MDP), although the difference can be as large as 3 kcal/mol. Finally, comparing the $\Delta G^{L \to H}$ values for the 25 combinations, we see that the results are mixed. On one hand, the results are largely consistent, for example, while the $\Delta G_0^{L \to H}$ values may differ by as much as 8 kcal/mol, once ΔG_Λ^L is included, the estimated $\Delta G^{L \to H}$ values are closer in magnitude with a standard deviation of 1.9 kcal/mol. On the other hand, for the serine case, one particular choice of $\mathbf{X}_0'$ leads to consistently different $\Delta G^{L \to H}$ values compared to the other 20 combinations by as much as ~6 kcal/mol. As shown in Fig. 10(d), the location of this reference structure (labeled 5) is not much different from the other four reference structures in the ϕ, ψ plane. On the other hand, the five reference structures differ in all degrees of freedom (they were collected from a fully unrestrained simulation), and projection in Fig. 10(d) reflects only two dihedral angles. Evidently, additional studies are required to develop better ways to determine optimal reference structures, such as using an efficient clustering analysis. Nevertheless, the current analysis highlights the importance of collecting multiple reference structures at both L and H levels to explicitly evaluate the robustness of the free energy results.

C. Length of sampling at the low- and high-level of theory

Finally, we briefly comment on the minimal amount of sampling needed at the L and H levels of theory for the test systems

FIG. 11. Convergence behavior of different free energy estimators for a blocked serine with a frozen structure in water with respect to the number of data points included. The L is CHARMM36, and the H is DFTB3/3OB/MM. Note that the reference M-λ TI value is different from that shown in Table I because a different serine structure is adopted in this set of calculations.

examined here. For sampling at the low level, we use the example of a frozen Ser in water, which exhibits the adequate distribution overlap between L = CHARMM36 and H = DFTB3/3OB/MM [Fig. 7(b)]. We examine the convergence of various free energy estimators for $\Delta G_0^{L \to H}$ with respect to the number of data points included. As shown in Fig. 11, even for this favorable case, at least 20 000 data

FIG. 12. Normalized dynamic Stokes shift computed for L = CHARMM36 and H = DFTB3/3OB/MM simulations. The four solution cases examined are (a) a fully unrestrained blocked serine, (b) a blocked serine with a frozen structure, (c) a fully unrestrained methyl diphosphate (MDP), and (d) a MDP with a frozen structure. The longer time constants from bi-exponential fits are 25.1 ps, 0.7 ps, 15.2 ps, and 1.8 ps for (a)–(d)], respectively. Similar results are found for L = DFTB3/3OB/MM and H = B3LYP/6-31G(d)/MM for the cases of frozen solutes.

points are necessary for good statistical convergence; this is in line with previous analysis.[20,25,34,53] As expected, the convergence behavior for BAR and LRA, which involve sampling at both L and H levels, is superior to the one-step FEP using either L or H sampling alone.

Another relevant question for multi-level free energy simulation is the time required to equilibrate at the H level, provided that adequate equilibration has already been done at the L level. For the computationally expensive H level, minimizing the amount of sampling is crucial to the practical efficiency. Here, we advocate the use of time-dependent Stokes shift, a quantity borrowed from the condensed phase spectroscopy literature.[81,82] The normalized dynamic Stokes shift is defined as

$$S(t) = \frac{\langle \Delta U(t) - \Delta U(\infty) \rangle}{\langle \Delta U(0) - \Delta U(\infty) \rangle}, \quad (14)$$

where $\Delta U(t)$ is the energy difference between the L/H levels at time t in the H level simulation that started (i.e., $t = 0$) with an equilibrated snapshot from the L level simulation. The quantity $S(t)$ reflects the time scale for the environment to respond to the change of the solute potential function from L to H.

Specifically, for the two test systems examined here, we examine the combination of L = CHARMM36 and H = DFTB3/3OB/MM and we compare the situation of free solutes and frozen solutes, which are expected to correspond to the limiting behaviors of the solvent response. To compute $S(t)$, we conduct 30 independent H level trajectories for each case examined.

As shown in Fig. 12, with a fully flexible solute, solvent response to the change of the solute potential function from CHARMM36 to DFTB3/3OB involves at least two time scales, and the longer time scale is on the order of 20 ps. Although MDP tends to exhibit a poorer distribution overlap between L and H as compared to Ser (see Fig. 7), the solvent response time is similar for the two solutes. With a frozen solute, the solvent response to the change of the solute potential function is much faster, even the longer time scale is merely a few picoseconds. These calculations suggest that with adequately equilibrated simulations at the L level for modest size solutes (QM regions), the amount of equilibration at the H level is likely limited to tens of picoseconds, which are readily affordable nowadays. Whether this is the case for a heterogeneous environment such as the active site of proteins remains to be examined in the future.

IV. CONCLUDING REMARKS

There has been tremendous interest in combining different levels of theory to obtain reliable thermodynamic properties for condensed phase systems. It has widely been recognized that the key to such multi-level free energy simulations is to ensure a favorable distribution overlap between different levels of theory. Therefore, the fundamental challenge is to identify the major degrees of freedom that lead to gaps in the distributions at the L and H levels of theory and then develop approaches that circumvent convergence difficulties due to the lack of adequate distribution overlap.

In this work, we have explored a strategy that converts from the L to H level of theory in a staged fashion, so as to ensure a favorable distribution overlap between different models along the

way. The key philosophy behind the strategy is to treat the problematic degrees of freedom (denoted as $\mathbf{X}$ in this work), which lead to distribution gaps, differently from the rest; the stiff degrees of freedom in $\mathbf{X}$ are treated with constraints (for bond lengths) or hard restraints (for bond angles), softer degrees of freedom (e.g., dihedrals) are treated with biasing potentials, and non-bonded degrees of freedom (e.g., partial charges) are treated by introducing an intermediate level of model (e.g., replacing MM charges with Mulliken/ESP charges).

Importantly, different models are connected through well-defined steps in a thermodynamic cycle, and corresponding free energy components are evaluated explicitly; this is an essential difference between our approach with more approximate schemes that compute, for example, only interaction energies. Test calculations using model compounds in solution indicate that the free energy components for connecting different levels of theory (e.g., ΔG_Λ^L for the $\mathbf{X}_0/\mathbf{X}_0'$ transformation) can be large in magnitude (~20 kcal/mol) and therefore essential to evaluate explicitly. While the staged transformation approach introduces additional steps, most of the free energy component calculations are done at the low level of theory and therefore do not significantly increase the computational cost. In fact, one way of describing our approach is starting with the L level conformational ensemble, progressively building up the optimal conformational ensemble that exhibits a favorable distribution overlap with the high-level of theory such that the amount of expensive H calculations is kept at a minimum.

As a proof of concept, two simple but non-trivial solution model systems are used to demonstrate that the staged transformation scheme is able to reproduce brute-force multi-window thermodynamic integration results for the L to H conversion with encouraging accuracy. For example, for a charged molecule (methyl diphosphate) in solution, free energy perturbation using only L level trajectories has a large error of ~13 kcal/mol, the staged transformation scheme leads to results within 1 kcal/mol–2 kcal/mol from M-λ TI calculations. We note that in the current implementation, we do make the approximation that confining the stiff degrees of freedom in $\mathbf{X}$ has similar free energy costs at the L and H levels of theory (i.e., $\Delta\Delta G_{cfn}^{L/H} \sim 0$); considering the localized nature of these degrees of freedom, the errors of the approximation are likely small, as supported by the favorable results for the staged transformation calculations, although this needs to be quantitatively evaluated for more complex systems in the future.

To apply the staged approach to more complex systems with a large number of atoms treated at multiple levels, the two major bottlenecks concern the systematic identification of the $\mathbf{X}$ degrees of freedom and the selection of appropriate reference structures at different levels of theory (i.e., $\mathbf{X}_0$, $\mathbf{X}_0'$). For the former issue, creative solutions likely involve combining multiple short simulations with machine learning algorithms for identifying variables that lead to large distribution gaps between different levels of theory, similar in spirit to the problem of optimizing collective variables in free energy simulations.[83] In fact, it is conceivable that machine learning approaches can be used to simultaneously learn the difference between the different levels of theory and the key coordinates responsible for the distribution gaps. In this way, the staged transformation approach and Δ-learning[51] can be integrated so that

convergence in $\Delta G^{L \to H}$ can be achieved with only modest improvement of the low-level method based on a limited amount of high-level data. For the issue of choosing reference structures, efficient clustering algorithms and dimensional reduction schemes should be explored and compared; test calculations illustrate the importance of sampling multiple reference structures to evaluate the robustness of the result. For cases that involve multiple free energy basins, it is natural to employ a divide-and-conquer strategy that requires evaluating free energy difference at the L/H levels of theory for the different basins separately.[84]

Finally, we stress the importance of conducting *some degree* of sampling at the H level of theory. This helps to explicitly evaluate the distribution overlap between the different models, and even a simple LRA model can be a substantial improvement over free energy perturbation that relies on the L level sampling alone, as shown by the model systems studied here. Along this line, monitoring the energy gap correlation function (i.e., the dynamic Stokes shift) can be an effective way to evaluate equilibration at the expensive H level, starting with a well-equilibrated L level ensemble.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol. **103**, 227–249 (1976).

[2] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations," J. Comput. Chem. **11**, 700–733 (1990).

[3] K. B. Lipkowitz and D. B. Boyd, in *Reviews in Computational Chemistry VII*, edited by J. Gao (VCH, New York, 1995), p. 119.

[4] G. Monard and K. M. Merz, "Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems," Acc. Chem. Res. **32**, 904–911 (1999).

[5] R. A. Friesner and V. Guallar, "*Ab initio* QM and QM/MM methods for studying enzyme catalysis," Annu. Rev. Phys. Chem. **56**, 389–427 (2005).

[6] D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, P. König, G. Li, D. Xu, H. Guo, M. Elstner, and Q. Cui, "Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes," J. Phys. Chem. B **110**, 6458–6469 (2006).

[7] H. M. Senn and W. Thiel, "QM/MM studies of enzymes," Curr. Opin. Chem. Biol. **11**, 182–187 (2007).

[8] H. Hu and W. Yang, "Free energies of chemical reactions in solution and in enzymes with *ab initio* quantum mechanics/molecular mechanics methods," Annu. Rev. Phys. Chem. **59**, 573–601 (2008).

[9] E. Brunk and U. Rothlisberger, "Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states," Chem. Rev. **115**, 6217–6263 (2015).

[10] J. Gao and D. G. Truhlar, "Quantum mechanical methods for enzyme kinetics," Annu. Rev. Phys. Chem. **53**, 467–505 (2002).

[11] J. Gao, S. Ma, D. T. Major, K. Nam, J. Pu, and D. G. Truhlar, "Mechanisms and free energies of enzymatic reactions," Chem. Rev. **106**, 3188–3209 (2006).

[12] J. Pu, J. Gao, and D. G. Truhlar, "Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions," Chem. Rev. **106**, 3140–3169 (2006).

[13] U. Ryde, "QM/MM calculations on proteins," in *Methods in Enzymology* (Academic Press, 2016), Vol. 577, pp. 119–158.

[14] S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, and W. Thiel, "P450 enzymes: Their structure, reactivity, and selectivity-modeled by QM/MM calculations," Chem. Rev. **110**, 949–1017 (2010).

[15] R. Lonsdale and A. Mulholland, "QM/MM modelling of drug-metabolizing enzymes," Curr. Top. Med. Chem. **14**, 1339–1347 (2014).

[16] N. Bernstein, J. R. Kermode, and G. Csányi, "Hybrid atomistic simulation methods for materials systems," Rep. Prog. Phys. **72**, 026501 (2009).

[17] F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H.-J. Werner, "High-accuracy computation of reaction barriers in enzymes," Angew. Chem., Int. Ed. **45**, 6856–6859 (2006).

[18] E. Rosta, M. Klähn, and A. Warshel, "Towards accurate *ab initio* QM/MM calculations of free-energy profiles of enzymatic reactions," J. Phys. Chem. B **110**, 2934–2941 (2006).

[19] J. Heimdal and U. Ryde, "Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations," Phys. Chem. Chem. Phys. **14**, 12592 (2012).

[20] X. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov, and Q. Cui, "QM/MM free energy simulations: Recent progress and challenges," Mol. Simul. **42**, 1056–1078 (2016).

[21] V. Luzhkov and A. Warshel, "Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies," J. Comput. Chem. **13**, 199–213 (1992).

[22] J. Gao, "Absolute free energy of solvation from Monte Carlo simulations using combined quantum and molecular mechanical potentials," J. Phys. Chem. **96**, 537–540 (1992).

[23] S. Martí, V. Moliner, and I. Tuñón, "Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems," J. Chem. Theory Comput. **1**, 1008–1016 (2005).

[24] M. Retegan, M. Martins-Costa, and M. F. Ruiz-López, "Free energy calculations using dual-level Born–Oppenheimer molecular dynamics," J. Chem. Phys. **133**, 064103 (2010).

[25] G. König, P. S. Hudson, S. Boresch, and H. L. Woodcock, "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 Comput. **10**, 1406–1419 (2014).

[26] I. Polyak, T. Benighaus, E. Boulanger, and W. Thiel, "Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation," J. Chem. Phys. **139**, 064105 (2013).

[27] P. Li, X. Jia, X. Pan, Y. Shao, and Y. Mei, "Accelerated computation of free energy profile at *ab initio* quantum mechanical/molecular mechanics accuracy via a semi-empirical reference potential. I. Weighted thermodynamics perturbation," J. Chem. Theory Comput. **14**, 5583–5596 (2018).

[28] P. Li, F. Liu, X. Jia, Y. Shao, W. Hu, J. Zheng, and Y. Mei, "Efficient computation of free energy surfaces of Diels–Alder reactions in explicit solvent at *ab initio* QM/MM level," Molecules **23**, 2487 (2018).

[29] M. Wang, Y. Mei, and U. Ryde, "Host-guest relative binding Affinities at density-functional theory level from semiempirical molecular dynamics simulations," J. Chem. Theory Comput. **15**, 2659–2671 (2019).

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

[31] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 826–843 (2011).

[32] A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free energy calculations," J. Phys. Chem. B **114**, 10235–10253 (2010).

[33] D. M. Zuckerman and T. B. Woolf, "Theory of a systematic computational error in free energy differences," Phys. Rev. Lett. **89**, 180602 (2002).

[34] U. Ryde, "How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging," J. Chem. Theory Comput. **13**, 5745–5752 (2017).

[35] R. H. Wood, W. C. F. Mühlbauer, and P. T. Thompson, "Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: Sample-size hysteresis," J. Phys. Chem. **95**, 6670–6675 (1991).

[36] D. A. Kofke, "Free energy methods in molecular simulation," Fluid Phase Equilib. **228-229**, 41–48 (2005).

[37] N. Lu, J. K. Singh, and D. A. Kofke, "Appropriate methods to combine forward and reverse free-energy perturbation averages," J. Chem. Phys. **118**, 2977–2984 (2003).

[38] D. Wu and D. A. Kofke, "Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation," J. Chem. Phys. **123**, 054103 (2005).

[39] D. Wu and D. A. Kofke, "Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations," J. Chem. Phys. **123**, 084109 (2005).

[40] D. A. Kofke, "On the sampling requirements for exponential-work free-energy calculations," Mol. Simul. **104**, 3701–3708 (2006).

[41] S. Boresch and H. L. Woodcock, "Convergence of single-step free energy perturbation," Mol. Phys. **115**, 1200–1213 (2017).

[42] M. Gaus, Q. Cui, and M. Elstner, "DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB)," J. Chem. Theory Comput. **7**, 931–948 (2011).

[43] M. Gaus, Q. Cui, and M. Elstner, "Density functional tight binding: Application to organic and biological molecules," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 49–61 (2014).

[44] F. Kearns, L. Warrensford, S. Boresch, and H. Woodcock, "The good, the bad, and the ugly: "HiPen", a new dataset for validating (S)QM/MM free energy simulations," Molecules **24**, 681 (2019).

[45] P. S. Hudson, S. Boresch, D. M. Rogers, and H. L. Woodcock, "Accelerating QM/MM free energy computations via intramolecular force matching," J. Chem. Theory Comput. **14**, 6327–6335 (2018).

[46] G. Hummer, "Fast-growth thermodynamic integration: Error and efficiency analysis," J. Chem. Phys. **114**, 7330–7337 (2001).

[47] C. Jarzynski, "Nonequilibrium equality for free energy differences," Phys. Rev. Lett. **78**, 2690–2693 (1997).

[48] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," Phys. Rev. E **60**, 2721–2726 (1999).

[49] I. Rossi and D. G. Truhlar, "Parameterization of NDDO wave-functions using genetic algorithms. An evolutionary approach to parameterizing potential-energy surfaces and direct dynamics calculations for organic-reactions," Chem. Phys. Lett. **233**, 231–236 (1995).

[50] Y. Zhou, P. Ojeda-May, M. Nagaraju, and J. Pu, "Toward determining ATPase mechanism in ABC transporters: Development of the reaction path-force matching QM/MM method," in *Methods in Enzymology* (Academic Press, 2016), Vol. 577, pp. 185–212.

[51] L. Shen and W. Yang, "Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks," J. Chem. Theory Comput. **14**, 1442–1455 (2018).

[52] N. V. Plotnikov, S. C. L. Kamerlin, and A. Warshel, "Paradynamics: An effective and reliable model for *ab initio* QM/MM free-energy calculations and related tasks," J. Phys. Chem. B **115**, 7950–7962 (2011).

[53] F. L. Kearns, P. S. Hudson, H. L. Woodcock, and S. Boresch, "Computing converged free energy differences between levels of theory via nonequilibrium

[54] G. König, B. R. Brooks, W. Thiel, and D. M. York, "On the convergence of multiscale free energy simulations," Mol. Simul. **44**, 1062–1081 (2018).

[55] T. J. Giese and D. M. York, "Development of a robust indirect approach for MM → QM free energy calculations that combines force-matched reference potential and Bennett's acceptance ratio methods," J. Chem. Theory Comput. **15**, 5543–5562 (2019).

[56] M. A. Olsson, P. Söderhjelm, and U. Ryde, "Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level," J. Comput. Chem. **37**, 1589–1600 (2016).

[57] P. S. Hudson, H. L. Woodcock, and S. Boresch, "Use of interaction energies in QM/MM free energy simulations," J. Chem. Theory Comput. **15**, 4632–4645 (2019).

[58] I. Andricioaei and J. E. Straub, "Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide," Phys. Rev. E **53**, R3055–R3058 (1996).

[59] T. J. Giese and D. M. York, "Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations," J. Chem. Phys. **127**, 194101 (2007).

[60] S. Kaminski, T. J. Giese, M. Gaus, D. M. York, and M. Elstner, "Extended polarization in 3rd-order SCC-DFTB from chemical potential equilization," J. Phys. Chem. A **116**, 9131–9141 (2012).

[61] A. S. Christensen, M. Elstner, and Q. Cui, "Improving intermolecular interactions in DFTB3 using extended polarization from chemical-potential equalization," J. Chem. Phys. **143**, 084123 (2015).

[62] T. J. Giese, H. Chen, T. Dissanayake, G. M. Giambaşu, H. Heldenbrand, M. Huang, E. R. Kuechler, T.-S. Lee, M. T. Panteva, B. K. Radak, and D. M. York, "A variational linear-scaling framework to build practical, efficient next-generation quantum force fields," J. Chem. Theory Comput. **9**, 1417–1427 (2013).

[63] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, "Semi-empirical quantum mechanical methods for non-covalent interactions for chemical and biochemical applications," Chem. Rev. **116**, 5301–5337 (2016).

[64] Q. Cui and M. Karplus, "Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations," J. Chem. Phys. **112**, 1133–1149 (2000).

[65] H. L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong, and B. R. Brooks, "Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit," J. Chem. Phys. **129**, 214109 (2008).

[66] G. König and B. R. Brooks, "Correcting for the free energy costs of bond or angle constraints in molecular dynamics simulations," Biochim. Biophys. Acta **1850**, 932–943 (2015).

[67] Y. Deng and B. Roux, "Computations of standard binding free energies with molecular dynamics simulations," J. Phys. Chem. B **113**, 2234–2246 (2009).

[68] Q. Cui, "Quantum mechanical methods in biochemistry and biophysics," J. Chem. Phys. **145**, 140901 (2016).

[69] C. H. Bennett, "Efficient estimation of free energy difference from Monte Carlo data," J. Comput. Phys. **22**, 245–268 (1976).

[70] M. R. Shirts and V. S. Pande, "Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration," J. Chem. Phys. **122**, 144107 (2005).

[71] S. Bruckner and S. Boresch, "Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method," J. Comput. Chem. **32**, 1303–1319 (2011).

[72] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, "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 Comput. **8**, 3257–3273 (2012).

[73] M. Gaus, A. Goez, and M. Elstner, "Parametrization and benchmark of DFTB3 for organic molecules," J. Chem. Theory Comput. **9**, 338–354 (2013).

[74] R. Stote, D. States, and M. Karplus, "On the treatment of electrostatic interactions in biololecular simulation," J. Chim. Phys. **88**, 2419–2433 (1991).

[75] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: The biomolecular simulation program," J. Comput. Chem. **30**, 1545–1614 (2009).

[76] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes," J. Comput. Phys. **23**, 327–341 (1977).

[77] N. Lu and D. A. Kofke, "Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling," J. Chem. Phys. **114**, 7303–7311 (2001).

[78] N. Lu and D. A. Kofke, "Accuracy of free-energy perturbation calculations in molecular simulation. II. Heuristics," J. Chem. Phys. **115**, 6866–6875 (2001).

[79] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," J. Comput. Chem. **31**, 671–690 (2010).

[80] F. Sittel and G. Stock, "Perspective: Identification of collective variables and metastable states of protein dynamics," J. Chem. Phys. **149**, 150901 (2018).

[81] R. M. Stratt and M. Maroncelli, "Nonreactive dynamics in solution: The emerging molecular view of solvation dynamics and vibrational relaxation," J. Phys. Chem. **100**, 12981–12996 (1996).

[82] W. H. Thompson, "Solvation dynamics and proton transfer in nanoconfined liquids," Annu. Rev. Phys. Chem. **62**, 599–619 (2011).

[83] Y. Wang, J. M. Lamim Ribeiro, and P. Tiwary, "Machine learning approaches for analyzing and enhancing molecular dynamics simulations," Curr. Opin. Struct. Biol. **61**, 139–145 (2020).

[84] A. Hodel, T. Simonson, R. O. Fox, and A. T. Brünger, "Conformational substrates and uncertainty in macromolecular free energy calculations," J. Phys. Chem. **97**, 3409–3417 (1993).