

BAR-based Multi-dimensional Nonequilibrium Pulling for Indirect Construction of QM/MM Free Energy Landscape: From Semi-empirical to Ab Initio

Zhaoxi Sun^{1,2*}

¹*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

²*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

*To whom correspondence should be addressed: z.sun@fz-juelich.de

Abstract

The indirect method for the construction of Quantum mechanics (QM)/ molecular mechanics (MM) free energy landscapes provides a cheaper alternative of free energy simulations at QM level. The indirect method features a direct calculation of the free energy profile at a relatively cheap low-level Hamiltonian and a low-level to high-level correction. In the thermodynamic cycle, the direct low-level calculation along the physically meaningful reaction coordinate is corrected via the alchemical method, which is often achieved with perturbation-based techniques. Often the indirect method can lead to about an order of magnitude speedup in free energy simulation. In our previous work, a multi-dimensional nonequilibrium pulling framework is proposed for the indirect construction of QM/MM free energy landscapes. The method relies on bidirectional nonequilibrium pulling and bidirectional reweighting with the statistically optimal estimator. In the previous work, we focus on obtaining semi-empirical QM (SQM) results indirectly from direct MM simulations and MM \leftrightarrow SQM corrections. In this work, we apply this method to obtain results under ab initio QM Hamiltonians by combining direct SQM results and SQM \leftrightarrow QM corrections. The indirect nonequilibrium scheme is tested on a dihedral flipping case and a series of SQM and QM Hamiltonians are benchmarked. It is observed that PM6 achieves the best performance among the low-level Hamiltonians, while AM1 and MNDO perform less well. Therefore, we recommend using PM6 as the low-level theory in the indirect free energy simulation. The comparison between the indirect results from different SQM Hamiltonians could also provide some hints on the development of charge models. As AM1 can be corrected with the bond charge correction (BCC) to provide a cheap and accurate charge model, which is able to accurately reproduce the electrostatic potential (ESP) at HF level, PM6 would be able to do

the same thing. Considering its higher similarity to the high-level Hamiltonians, the PM6-BCC model could be more accurate than the existing AM1-BCC model. Another central result in the current work is a basic protocol of choosing the strength of restraints and an appropriate time step in nonequilibrium free energy simulation at the stiff spring limit. We provide theoretical derivations to emphasize the importance of using a sufficiently large force constant and choosing an appropriate time step. It is worth noting that a general rule of thumb for choosing the time step, according to our derivation, is that a time step of 1 fs or smaller should be used, as long as the stiff spring approximation is employed, even in simulations with constraints on bonds involving hydrogen atoms.

I. Introduction

Free energy differences between different states of a system or different systems are the thermodynamic determinant of the tendency of physical processes. The variation of free energy along generalized coordinates is given by the so-called free energy profile or free energy landscape. As the timescale of processes of interest is often longer than molecular dynamics (MD) accessible timescales, obtaining converged statistics from brute-force simulations is impossible. To overcome this timescale issue, smart sampling regimes are devised. Enhanced sampling methods are powerful tools to accelerate the sampling and construct the free energy profiles or surfaces. Generally, they can construct the free energy surfaces with only a fraction of computational cost of the conventional MD simulations.¹⁻⁴ Enhanced sampling techniques add biasing potentials to enhance the sampling efficiency in specific regions in the phase space. An example in point is umbrella sampling,^{3, 5-7} where harmonic biasing potentials are added. Post-processing is needed to reweight the samples obtained from the biased ensembles to recover the unbiased statistics in the original unperturbed system. Most of the popular reweighting techniques are based on free energy perturbation (FEP). For instance, the popular multi-state reweighting estimators of WHAM⁸ are the multi-state version of FEP. As FEP suffers from sample size hysteresis,⁹⁻¹¹ its bidirectional variant named Bennett Acceptance Ratio (BAR) is often used to achieve the best statistical efficiency.¹²⁻¹³ The multi-state version of BAR is MBAR,¹⁴ which makes the best use of samples from multiple equilibrium states. vFEP¹⁵ is another popular perturbation-based reweighting method, where curve fitting is also incorporated in the construction of the free energy profile. Recently, the nonequilibrium pulling method named Steered MD (SMD) is coupled to the nonequilibrium generalization of FEP and BAR, i.e. Jarzynski's Identity (JI)¹⁶ and Crooks' Equation (CE),¹⁷ providing a nonequilibrium alternative for the construction of the potential of mean force (PMF).^{9-11, 18-23}

As a state function, the free energy difference between different states can, in principle, be calculated with any pathway connecting these states. Practical considerations on the efficiency of the pathway are often technical. The identification of the optimal pathway can be performed by minimizing the thermodynamic length along it.²⁴ Adaptive sampling along the whole pathway defined by the generalized coordinate named collective variable (CV) achieves the similar goal, i.e. minimizing the statistical uncertainty.²⁵⁻²⁷ A physical process happens in a physical space and can be described with a set of physically meaningful CV. PMF along these coordinates depicts the variation of thermodynamics in the process. If a non-physical pathway connecting two physical systems is defined, the free energy calculation is performed to calculate the free energy profile along this non-physical pathway, which is often termed as the alchemical method.²⁸⁻³² As the

data are obtained in the biased and non-physical ensemble, still the reweighting procedure is needed to calculate the generalized PMF along the pathway. As the perturbation method relies on sufficient phase space overlap to interpolate/extrapolate the data, similar to the introduction of a series of umbrella windows to ensure overlap between neighboring windows, a series of intermediate states described with artificial Hamiltonians are often inserted between the two physical states at the terminals of the transformation. The alchemical method can be very efficient in the calculation of free energy difference between equilibrium states. However, as the free energy profile is constructed in the alchemical space, no physical picture of the thermodynamics in physically meaningful transformations such as conformational changes in biomolecular systems can be obtained. In order to investigate the behavior of the system in the physical world, the free energy landscape in the physically meaningful space is still required.

Enhanced sampling techniques can be used to obtain a converged picture of the thermodynamics of the system. However, the accuracy of the computer simulation is determined by the description of the system. The energetics of the system are provided by the Hamiltonian. Quantum mechanics (QM) Hamiltonians are more accurate but expensive,³³⁻³⁵ while molecular mechanics (MM) or even coarser models are computationally efficient but less accurate. Even with advanced computational resources, the full QM treatment of a large system is still prohibitive. Multiscale models are proposed to extend the applicability of QM Hamiltonians. For instance, QM/ semi-empirical QM (SQM), QM/MM and fragmentation methods are able to provide an accurate description of large systems.^{33, 36-39} However, these calculations are still computationally demanding for converged phase space sampling.

The indirect QM/MM free energy simulation technique, which combines the free energy simulation along physical CVs at a cheaper yet good Hamiltonian and an alchemical free energy calculation to connect the cheaper and the computationally demanding Hamiltonians, is proposed to provide an alternative way to obtain the results in direct QM/MM free energy simulations.⁴⁰⁻⁴⁷ A thermodynamic cycle depicted in Fig. 1a is constructed. The direct free energy simulation under the high-level Hamiltonian is obtained from the direct free energy simulation under the low-level Hamiltonian and a low-level-to-high-level correction.⁴⁰⁻⁴⁹ Here, the low level of theory is a Hamiltonian with lower computational cost and accuracy, while the high-level one is that with higher computational burden and accuracy. The low-level ones are often chosen as MM or SQM Hamiltonians, while the high-level ones could be SQM or ab initio QM Hamiltonians. Such an indirect transformation pathway is generally about an order of magnitude faster than the direct one and the speedup is system-dependent.^{45, 50-61} Published studies mostly relied on equilibrium techniques in the direct low-level free energy simulations and the low-to-high correction. Only a few studies employed

nonequilibrium techniques.²² In our previous work, we report a general nonequilibrium framework for the indirect construction of QM/MM free energy landscapes, where all transformations are performed in a nonequilibrium way. The nonequilibrium stratification or the staging nonequilibrium pulling is coupled to the statistically optimal bidirectional reweighting estimator of BAR/CE in both the alchemical²⁶ space and the configurational one.²⁷ The resulting two-dimensional (2D) nonequilibrium free energy simulation provides an indirect way to construct the QM/MM free energy landscapes. The Savitzky-Golay filter is employed to increase the signal-to-noise ratio and get a smoother PMF. The computational framework is applied to one periodic dihedral case of solvated ACE-NME and one non-periodic distance case of the stretching of deca-alanine. In our previous work, the low-level Hamiltonian is set to MM, while the high-level one is SQM. In the current work, we extend the protocol by setting the low-level Hamiltonian as SQM and use ab initio QM Hamiltonians as the high-level one. Three SQM Hamiltonians including the Austin Model 1 (AM1),⁶² the Parametrized Model number 6 (PM6)⁶³ and the Modified Neglect of Diatomic Overlap (MNDO)⁶⁴ are tested. As for the ab initio QM Hamiltonians, the Hartree-Fock method (HF),⁶⁵⁻⁶⁷ the post-HF method of second-order perturbation theory (MP2),⁶⁸⁻⁷¹ the Becke 3-parameter Lee–Yang–Parr (B3LYP) functional,⁷²⁻⁷⁴ and the ω B97X-D functional⁷⁵ are employed. We expect this benchmark test to provide some hints on the selection of the low-level Hamiltonian to achieve better efficiency. Another inspiring point obtained from this benchmark test is that PM6 may be a nice candidate to generate the atomic charges for MD simulations, due to its similarity to the high-level Hamiltonians.

The stiff spring limit is often used in nonequilibrium pulling. Such an approximation requires a sufficiently large force constant to restrain the system close enough to the desired state. Although some discussions about the width of the distribution of CV and the selection of force constants have been presented,⁷⁶⁻⁷⁷ nothing is discussed about the selection of the time step used in the biased simulation. However, from a theoretical perspective, with a large force constant to achieve the stiff spring limit, the time step should also be adjusted accordingly. Therefore, we revisit this width-of-distribution issue and then extend the derivation to the acceptable range of time steps. We emphasize the use of a small time step to integrate the equation of motions, and recommend that 1 fs or smaller should be used in simulations even with constraints on chemical bonds (involving hydrogen atoms).

II. Methodology

1. A brief review of the multi-dimensional nonequilibrium pulling scheme for the indirect construction of free energy landscapes at QM level.

The calculation of the free energy landscape at the high-level theory is decomposed to a 2D enhanced sampling along one configurational CV and one alchemical CV. The SMD simulations along the configurational CV is achieved by adding a time-dependent harmonic potential defined by Eq. (1), which drives the system from one state to another.⁷⁸

$$V(\mathbf{q}) = \frac{k}{2} (\xi(\mathbf{q}) - \xi_0(t))^2 \quad (1)$$

, where V is the harmonic potential, k denotes the force constant, \mathbf{q} represents the coordinate vector, ξ_0 is the time-dependent CV defining the configurational transformation, and ξ refers to the current value of the CV. Such a restraint restrains the CV to the desired value, but there are still fluctuations in the value of CV. With a large enough force constant, the fluctuation of CV is sufficiently suppressed, which is often called as the stiff spring limit.⁷⁹⁻⁸¹ In alchemical space, the generalized CV of the alchemical order parameter moves in the same way. The alchemical CV varied with the pre-defined schedule but without fluctuations in its value.

A long nonequilibrium pulling suffers from huge dissipations. The resulting convergence of the simulation is very slow and a huge number of samples are needed. Further, a long waiting time is required before we can get some feedback from simulations, which is a severe drawback in testing a protocol. To enhance the convergence, we need either to lengthen the time for pulling (i.e. use a slower pulling speed) or cut the whole transformation into smaller segments. In the nonequilibrium stratification or the staging nonequilibrium pulling scheme, we consider the case that the whole transformation is divided into K states and there are $K - 1$ segments.

The best way to reweight data from multiple equilibrium states is bidirectional reweighting with BAR/CE or multi-state reweighting with WHAM, MBAR or vFEP. As the reliability and the weighting factor are closely related to the overlap between different states, it is often observed that the results of perturbations are mostly determined by data from neighboring windows. As a result, bidirectional reweighting often provides very similar results to those obtained from multi-state reweighting methods. In nonequilibrium simulations, the transformation is essentially pair-wise. Namely, only nonequilibrium pulling between neighboring windows is performed. Therefore, the data obtained from nonequilibrium simulations should be analyzed with the bidirectional reweighting estimator of CE to achieve the best statistical efficiency. The statistically optimal and asymptotically unbiased two-state free energy estimator of CE is described with Eq. (2).^{12-13, 17}

$$\left\{ \begin{array}{l} \Delta A_{ij} = \ln \frac{\langle f(W_{ji} + C) \rangle_j}{\langle f(W_{ij} - C) \rangle_i} + C \\ C = \Delta A_{ij} + \ln \left(\frac{n_j}{n_i} \right) \end{array} \right. \quad (2)$$

, where A is the dimensionless free energy, ΔA is thus the corresponding difference between free energies of different states, $\langle \dots \rangle_i$ represents to the canonical average over nonequilibrium realizations initiated from state i , W_{ij} is the dimensionless work accumulated during the nonequilibrium pulling initiated from state i and ended in state j , n denotes the number of samples in each ensemble, and f is the Fermi function. The bidirectional estimator is very robust and the convergence is much faster than the unidirectional estimators of FEP and JI. If the distribution of work follows some well-defined models (e.g. Gaussian distribution), unidirectional estimators can be very efficient. However, if the distribution does not satisfy such approximation, unidirectional estimates introduce significant systematic errors. The magnitude of the systematic error is unknown, which makes it hard to judge the quality of the estimate. The general form of CE is able to get unbiased estimates with a reasonable sample size, which makes it generally applicable to all kinds of distributions. Therefore, we focus on the general formula of Eq. (2) rather than using parametric approximations.

The dimensionless variance of the CE estimates can be expressed as

$$\sigma_{ij}^2 = \frac{\text{Var}(f_{ij})}{n_i^2 f_{ij}^2} + \frac{\text{Var}(f_{ji})}{n_j^2 f_{ji}^2} \quad (3)$$

, where Var represents the absolute variance. The definition of f_{ij} is given by

$$\begin{aligned} f_{ij} &= \langle f(W_{ij} + C_{ij}) \rangle_i \\ f_{ji} &= \langle f(W_{ji} - C_{ij}) \rangle_j \\ n_{ij} f_{ij} &= n_{ji} f_{ji} \text{ from SCF of CE} \end{aligned} \quad (4)$$

, where j is larger than i and is often set to $j=i+1$, as only the transformations between neighboring intermediate states are performed. The requirement of using Eq. (3) to compute the statistical error of the free energy difference is that the samples should be uncorrelated. To extract independent initial configurations for nonequilibrium realizations from equilibrium ensembles, initial configurational sampling is needed. In this step, the autocorrelation function of the reaction coordinate we bias is calculated, and the statistical inefficiency ϕ_i is calculated with $\phi_i = 1 + 2\tau_i$, where τ_i is the autocorrelation time in state i .

Subsampling the whole dataset of initial configurations by the statistical inefficiency, we obtain a new set of independent configurations.

As the statistical inefficiency can be used to estimate the computational cost of getting an independent sample from equilibrium sampling, we likewise determine the computational cost in nonequilibrium pulling. We rewrite the statistical inefficiency in equilibrium dynamics as $\phi_{\text{eq},i}$, and add the length of each nonequilibrium pulling simulation $\phi_{\text{NEW},i}$. Therefore, the overall computational cost for each independent point in nonequilibrium pulling is

$$\phi_i = \phi_{\text{NEW},i} + \phi_{\text{eq},i} \quad (5)$$

. The pulling simulation is performed bidirectionally, i.e. in the forward (ξ increasing) and backward (ξ decreasing) directions. Therefore, $\phi_{\text{NEW},i}$ should be multiplied by 2 in the above equation. Further, to compare the computational cost of the direct and indirect schemes, we also need to scale the computational cost of each independent data point by the ratio of the computational costs under different Hamiltonians. In the current case, as the SQM calculation is much faster than the QM one, we scale all the computational cost to the SQM level, which enables a direct comparison of computational costs of different free energy simulation schemes. The resulting statistics are provided in Table 1, from which we know that the simulations at different SQM levels are very similar in efficiency, while the computational cost of ab initio QM calculations shows obvious dependence on the level of theory.

After obtaining the free energy difference between neighboring states, we accumulate the free energy difference in these segments to get the variation of free energy in the whole process,

$$\Delta A_{1k} = \sum_{i=1}^{k-1} \Delta A_{i,i+1} \quad (6)$$

$$\sigma_{1k}^2 = \sum_{i=1}^{k-1} \sigma_{i,i+1}^2 = \sum_{i=1}^{k-1} \left(\frac{\text{Var}(f_{i,i+1})}{n_i^2 f_{i,i+1}^2} + \frac{\text{Var}(f_{i+1,i})}{n_{i+1}^2 f_{i+1,i}^2} \right) \quad (7)$$

. If we set $k = K$, Eq. (6) gives the overall free energy difference along the pathway. By setting the free energy of state 1 as zero, i.e. the reference point

$$A_1 = 0 \quad (8)$$

, we obtain the free energy landscape by adding Eq. (6) to it,

$$A_k = \Delta A_{1k} = \sum_{i=1}^{k-1} \Delta A_{i,i+1} \quad (9)$$

. The above equations enable us to get the free energy profile in one dimension. By constructing a similar one-dimensional free energy profile along a CV orthogonal to the first CV, we obtain the multi-dimensional free energy surface. We define the first CV as the configurational CV with K_1 states and the second CV as the alchemical CV with K_2 states. Then, the free energy of the state (k_1, k_2) in the 2D free energy simulation is expressed as,

$$A_{k_1 k_2} = \Delta A_{k_1, 1, 11} + \Delta A_{k_1 k_2, k_1 1} = \sum_{i=1}^{k_1-1} \Delta A_{i, 1, i+1, 1} + \sum_{j=1}^{k_2-1} \Delta A_{k_1, j, k_1, j+1} \quad (10)$$

, where $\Delta A_{k_1, 1, 11}$ represents the free energy difference between configurational state 1 and configurational state k_1 at the 1st alchemical state. The statistical error in this case can be propagated with the normal error propagation procedure. For the same system, the descriptions under different Hamiltonians are often similar. Therefore, in the SQM \leftrightarrow QM correction, we do not employ the staging regime but do it in a single-step pulling. We define $K_2 = 1$ as the low-level Hamiltonian and $K_2 = 2$ as the high-level theory. Thus, bidirectional pulling is performed between $(k_1, 1)$ and $(k_1, 2)$. Then, the indirect scheme is performed by constructing the free energy profile along K_1 at the $K_2 = 1$ alchemical state and performing $(k_1, 1) \leftrightarrow (k_1, 2)$ corrections at each k_1 .

Convergence check in nonequilibrium simulations. There are two things needed to be checked in nonequilibrium pulling. The first one is the sample-size dependence of various ensemble averages, and the second one is the pulling-speed dependence.^{25-26, 82} As our central goal is the free energy, we focus on the time-evolution of the free energy profile and the comparison between free energy profiles obtained under different pulling speeds. As the statistical error or uncertainty is often more biased than the free energy itself, these variances are also needed to be checked.^{18, 25-26, 82} Convergence is reached when identical estimates of the free energy profiles are obtained at different pulling speeds and with different sample sizes, and the standard error is small enough and it decreases monotonically with the sample size. We note that this check needs to be performed for all segments in multi-stage nonequilibrium pulling.

In the calculation of the speedup of the indirect scheme, for simplicity, we use the equal sample size rule of $n_i = n_j = n$. This provides a quick estimate of the efficiency of the method, but may lead to some underestimation of the efficiency. More detailed and complicated calculations may be useful in this aspect, but this is good enough for our illustrative calculation.

2. Achieving the stiff spring limit.

In umbrella-sampling-related simulations, harmonic potential energy functions are added to the potential energy surface (PES) to enhance the sampling. Mechanical observables are selected to describe important motions and distinguish different states of the system. To initiate a simulation, we need to try how large a force constant should be used.

When a harmonic potential with force constant k ,

$$V = \frac{k}{2}(x - \mu)^2 \quad (11)$$

, is added to the flat PES V_0 with $\frac{\partial V_0}{\partial x} = 0$, where x is the value of the CV and μ is the center of the umbrella potential (i.e. the desired value of CV), the distribution of CV can be obtained by

$$V + V_0 = V_{tot} = -\frac{1}{\beta} \ln P \quad (12)$$

, where β is the reciprocal temperature, and P is the probability distribution of CV. Here, we keep the temperature factor rather than using the dimensionless energy, as the width of CV distribution and the selection of time step to integrate the equation of motions are temperature-dependent. Under this potential, the distribution of CV is Gaussian, namely

$$P = e^{-\beta V_0} e^{-\beta V} = e^{-\beta V_0} e^{-\frac{\beta k}{2}(x - \mu)^2} \quad (13)$$

. We can see that the exponential of V_0 can be transformed to the normalization factor for the Gaussian distribution. The probability density for an arbitrary normal distribution is

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \quad (14)$$

, where μ denotes the center of the distribution and σ is the standard deviation. Comparing Eq. (13) and Eq. (14), we have

$$\beta k = \frac{1}{\sigma^2} \quad (15)$$

. Note that the non-zero $\frac{\partial V_0}{\partial x}$ could lead to anharmonicity in the above distribution.

Practical value of k in biased simulations. At 300 K, β is $(0.59 \text{ kcal}/(\text{mol}))^{-1}$. We hereafter discuss the use of the force constant from a practical perspective, under the realization of a commonly used software for biasing the simulations, i.e. AMBER used in this work, coupled to a realization of the post-processing

reweighting algorithms, including WHAM,⁸ MBAR¹⁴ and vFEP¹⁵.

Take the definition of force constant in AMBER⁸³ as an example

$$V = k_{\text{AMBER}} (x - \mu)^2 \quad (16)$$

. Note that in software including WHAM,⁸ MBAR¹⁴ and vFEP,¹⁵ the definition of the force constant is the conventional definition provided by Eq. (11). Hence, we need to multiply k_{AMBER} by a factor of 2 in using the post-processing software for reweighting.

If the CV is an angle or a dihedral, in amber the unit of k is kcal/(mol*rad²), while the software use force constants measured in degrees. Therefore, the force constant for dihedral includes also a unit conversion. The resulting force constant in Eq. (11) should be

$$k = k_{\text{AMBER}} * \left(\frac{2\pi}{360}\right)^2 * 2 = k_{\text{AMBER}} * 0.0006092 \quad (17)$$

. If we set k as 200 kcal/(mol*rad²) the theoretical σ is 2.2°. If k equals 500 kcal/(mol*rad²), σ becomes 1.4°. As in practical cases the PES is not flat, the standard deviation would deviate from theoretical values.

Scanning k we have Fig. S1a. We start from 50 kcal/(mol*rad²) as the normal range we use is from 100 kcal/(mol*rad²) to 10000 kcal/(mol*rad²). Too small value leads to brute force simulation.

As in umbrella sampling the distributions of CV in neighboring windows should overlap, we need to choose a not-so-large force constant. For instance, for 5° increments, the usable k_{AMBER} is 100-200 kcal/(mol*rad²) and for 2° increments it is 500 kcal/(mol*rad²), in order to take a balance between sufficient phase space overlap and enhancing the sampling of CV.

If the aim is to achieve the stiff spring limit, in the case of nonequilibrium pulling and the initial configurational sampling of nonequilibrium free energy simulation to extract independent configurations, a large force constant is needed to restrain the CV sufficiently close to the desired value. With the largest force constant we test, i.e. 10000 kcal/(mol*rad²), σ is about 0.3°. Obviously, the stiff spring limit is well-achieved. The deviation from the desired value is always smaller than 1°, which is about 3σ for $\sigma = 0.3^\circ$. As a result, such a large force constant is often used to get initial configurations for nonequilibrium free energy simulations.^{27, 84-85} In the normal case, we recommend a force constant of 2000 kcal/(mol*rad²) (width of distribution $\sigma \sim 0.7^\circ$), which is already good enough to ensure a precise definition of the dihedral state. Also scanning the temperature space leads to Fig. S1b. We can see in the normal temperature range, the conclusion does not change much with the 300 K one.

If the CV is a distance, in amber the unit of k is kcal/(mol*Å²). Only multiplying it by 2 gives the conventional force constant in Eq. (11)

$$k = k_{\text{AMBER}} * 2 \quad (18)$$

. Scanning k in this case, we have Fig. S2a.

In this situation, in umbrella sampling simulations, 20 kcal/(mol*Å²) is usable for 0.2 Å increments to balance the overlap and the sampling enhancement, if there is no hard wall such as vdW walls. Too small values of the force constant may lead to shifted distribution caused by large $\frac{\partial V_0}{\partial x}$, the result of which is that the specific areas are not sampled well.

As for achieving the stiff spring limit, a force constant of 100 kcal/(mol*Å²) is already good enough to restrain the system within about 0.05 Å of the center of the biasing potential. The application of such a force constant to achieve the stiff spring limit can be seen in references.^{27, 85} With a huge force constant of about 2000 kcal/(mol*Å²), the fluctuation of CV is about 0.03 Å. The resulting smallest distance increment in nonequilibrium free energy simulation could be as small as 0.1 Å.

The overall dependence on temperature and the force constant is shown in Fig. S2b. The plot should be similar to Fig. S1b as they are only scaled by a constant. Some discussions about the width of distribution and the selection of the force constant can be seen in references.⁷⁶⁻⁷⁷

Choosing an appropriate time step. With a suitable force constant to bias the simulation, we then need to define how long the time step should/could be in the integration of the equation of motions. The accuracy and stability of the dynamics depend on the integration time step (ITS). The smaller the ITS is, the higher the accuracy is achieved. However, the computational cost increases with the magnitude of ITS. Therefore, finding an appropriate value of ITS is important in numerical simulations. The maximum ITS is often chosen as the minimum length of time step being able to resolve the fastest motion in the system.

A general rule of thumb of choosing an appropriate length of ITS is using values ranging from $\frac{T_{\min}}{20}$ to $\frac{T_{\min}}{5}$, where T_{\min} represents the period of the fastest motion in the system. These values ensure a reasonable precision in the numerical integration and are computationally efficient. Note that the lower bound of $\frac{T_{\min}}{20}$ is often recommended due to accuracy consideration and the upper bound is often not used.

Values smaller than the lower bound of $\frac{T_{\min}}{20}$ such as $\frac{T_{\min}}{30}$ or even $\frac{T_{\min}}{170}$ (for significant instantaneous perturbations, i.e. stepped loads to the system) are sometimes required.

Normally, in MD simulation, with bond-length-constraint algorithms such as SHAKE,⁸⁶⁻⁸⁸ we can use

an ITS as long as 2 fs. The application of the biasing potential or force may lead to a smaller T than the motions involving hydrogen atoms and lowers the maximum value applicable for ITS. Below, we provide discussions about the time step acceptable in MD simulation with the biasing potentials.

We still use the spring model and assume the effective mass of the atom linked by the spring as m g/mol. As we know, the period of a spring can be estimated by

$$T = 2\pi\sqrt{\frac{m}{k}} \quad (19)$$

. The effective mass $m = \frac{m_1 m_2}{m_1 + m_2}$, where m_1 and m_2 are the mass of the two atoms or groups,

respectively. Note that the harmonic mean of the masses is mainly determined by the smaller one. If one of the masses is much smaller than the other, e.g. in the case of C-H bonds, the effective mass is approximately the mass of the hydrogen atom. Therefore, in the following discussion, we use the smaller mass to approximate the effective mass, which will not lead to significantly different conclusions compared with the situation using the theoretically rigorous definition of the effective mass.

Again, we consider the dihedral CV in the first place. As k in amber is in unit of kcal/(mol*rad²), which is already in the SI (international system of units) base unit, we do not need to transform it with Eq. (17) to estimate the period. Transforming k and m into SI unit, we have

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{\tilde{m}/1000}{\tilde{k} * 4185 / r^2} \frac{kg / mol}{J / (mol * rad^2) / (m / rad)^2}} = 0.00307r\sqrt{\frac{\tilde{m}}{\tilde{k}}} s / m \quad (20)$$

. Here, we apply the transformation of $V = k(\Delta\phi)^2 = \frac{k}{r^2}(\Delta\phi \cdot r)^2$ to transform the displacement from angle $\Delta\phi$ to distance $\Delta\phi \cdot r$ for dimension reduction. The normal length of covalent bonds (radius r) is about 1.5 Å. \tilde{m} and \tilde{k} are the dimensionless mass and the dimensionless force constant, respectively. Note that the base unit of mass is kg rather than g, although there is a multiplier prefix of kilo-. Thus, a factor of 1000 is divided in the above equation. Also, we assume that 3 atoms of 4 atoms for defining the dihedral or 2 atoms of 3 atoms defining the angle remain frozen and only the other atom travels. Transforming the unit to femtosecond, we have

$$T = 0.00307r\sqrt{\frac{\tilde{m}}{\tilde{k}}} s / m = 461\sqrt{\frac{\tilde{m}}{\tilde{k}}} fs \quad (21)$$

. For hydrogen atom with mass 1 g/mol and an approximated force constant of 200 kcal/(mol*rad²) (meaning $k_{\text{AMBER}} = 100 \text{ kcal} / (\text{mol} \cdot \text{rad}^2)$), the acceptable time-step range is

$$1.6 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 7 \text{ fs} \quad (22)$$

. Here, $dt = \Delta t$ represents the time step. We can see that the lower bound is between the recommended time steps of 1 fs and 2 fs. Therefore, we need to use a smaller 1 fs time step in simulations. However, in biased simulations, we often do not add biasing potentials on hydrogen atoms but only include heavy atoms in the CV. For heavy atoms such as O or C atoms with masses of approximately 10 g/mol, under the force constant of 200 kcal/(mol*rad²), the acceptable time-step range is

$$5 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 21 \text{ fs} \quad (23)$$

. The lower bound is larger than the recommended ITS of 1 fs or 2 fs. Therefore, in umbrella sampling with small force constants, we do not need to change the ITS in the biased simulation. However, in the case of using a large force constant of 20000 kcal/(mol*Å²) (meaning $k_{\text{AMBER}} = 10000 \text{ kcal} / (\text{mol} \cdot \text{rad}^2)$) to achieve the stiff spring limit, the acceptable ITS is

$$0.5 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 2.1 \text{ fs} \quad (24)$$

. In this case, a smaller ITS of 1 fs or 0.5 fs is required. We note that if the dihedral is defined by groups of atoms,^{84, 89} the radius r and the effective mass m should be larger. As a result, the force on each atom is smaller and a longer ITS can be used.

When applying the above procedure to the distance CV, as k is in unit of kcal/(mol*Å²), we transform k and m into the SI base unit and get

$$T = 2\pi \sqrt{\frac{m}{k}} = 2\pi \sqrt{\frac{m/1000}{k * 4185 * 10^{20}} \frac{\text{kg} / \text{mol}}{\text{J} / \text{mol} / \text{m}^2}} = 307.1 \sqrt{\frac{\tilde{m}}{\tilde{k}}} \text{ fs} \quad (25)$$

. For hydrogen atom with the mass of 1 g/mol and an approximated force constant of 100 kcal/(mol*Å²), the range of ITS should be

$$1.5 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 6 \text{ fs} \quad (26)$$

. As the force constant of chemical bonds in force fields is about 100 kcal/(mol*Å²) and the effective mass for heavy-atom-hydrogen bonds is very close to the mass of hydrogen atoms, the above equation gives an explanation of why without SHAKE we always use a smaller time step of 1 fs. In the biased simulations, we are often restraining heavy atoms such as O and C and their masses are approximately 10 g/mol. Then for a force constant of 100 kcal/(mol*Å²)

$$5 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 19 \text{ fs} \quad (27)$$

, which would not change $\frac{T_{\text{min}}}{20}$ of the system. For a huge force constant of 2000 kcal/(mol*Å²) (meaning

$k_{\text{AMBER}} = 1000 \text{ kcal} / (\text{mol} \cdot \text{\AA}^2)$), the acceptable time step follows

$$1 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 4 \text{ fs} \quad (28)$$

. A practical example is the distance between the sodium and chloride ions in solvated Na-Cl ion pair,²⁷ the effective mass is about 20 g/mol. The largest force constant used in our previous work is 1500 kcal/(mol*Å²), which leads to the acceptable range for ITS of

$$1.7 \text{ fs} \approx \frac{T}{20} \leq dt \leq \frac{T}{5} \approx 7 \text{ fs} \quad (29)$$

. Hence, we can see that when trying to use the stiff spring approximation, a time step of 0.5 fs or 1 fs should be used. For not-so-large force constants used in umbrella sampling simulations, the time step of 1 fs or even 2 fs is already safe and good enough.

We scan the acceptable time steps and the results are given in Fig. 2a and Fig. 2b. In between the black ($\frac{T}{5}$) and red ($\frac{T}{20}$) curves are time steps usable. We should note that in the discussion about the time step dt above, only the harmonic restraining potential is considered. For numerical precision, values a little smaller than the $\frac{T}{20}$ bound are recommended. We can see that in the normal ranges of harmonic potentials applied in the simulations to achieve the stiff spring limit, the $\frac{T}{20}$ curve is often higher than 1 fs (the recommended value of ITS without the SHAKE constraints) but lower than 2 fs (the recommended value of ITS with SHAKE). Therefore, due to stability considerations, we recommend using an ITS of 1 fs in nonequilibrium pulling simulations with time-(in)dependent harmonic biasing potentials. Smaller values such as 0.5 fs are also applicable when one wants to achieve the stiff spring limit and uses a large force constant. Values larger than 1 fs may lead to numerical instabilities and unexpected errors in the system. For smaller force constants in the case of umbrella sampling, a time step of 2 fs is still usable.

A note we would like to add here is that the discussion above is based on the simulation protocol that a single ITS is used. By applying the multiple time step techniques,⁹⁰⁻⁹² we do not need to adjust the ITS for the whole system. Another useful discussion about the stiffness of the spring can be found in references.⁷⁷

3. Our protocols for the indirect free energy simulation under the nonequilibrium framework.

According to the above theoretical derivation and discussion, the following protocols are what we stick to:

- a) Use nonequilibrium pulling to achieve a larger phase space overlap and thus enhance convergence.
- b) Use bidirectional reweighting to achieve the best statistical efficiency.
- c) Do not use parametric approximations, e.g. Gaussian approximation, to achieve the best generality.

- d) Use independent samples to avoid underestimating the statistical uncertainty.
- e) In calculating the computational cost of each method, use the equal sample size rule to avoid exaggerating the performance of the multi-dimensional nonequilibrium method.
- f) As for the stiff spring limit, use a sufficiently large force constant to minimize the fluctuation of CV and choose an appropriate value of ITS.

The protocol a is the motivation of the current work. Although this may lead to some loss of computational efficiency if a single-step perturbation is already small enough for reliable convergence, with this nonequilibrium framework, the convergence can always be achieved by using a slower pulling speed without any change of the theoretical framework. Note that the single-step perturbation can also be achieved in this multi-dimensional reweighting framework by simply replacing the nonequilibrium work with the energy difference. Further, the number of samples needed for convergence is also significantly decreased upon using a slower pulling speed. Another merit of the nonequilibrium technique is the efficiency of parallelization. As each nonequilibrium pulling step lasts only for several hundred fs, each job can finish within a reasonable time with only a single core, which makes the best use of the available computational resources. The resulting performance of the simulation is thus very close to the benchmark test shown in Table 1. The protocols b and c provide the statistically optimal and the generally applicable way to deal with data from almost all complex systems, which is achieved by Eq. (2). The protocol d makes the calculation of statistical error theoretically meaningful, which is the requirement for data as input for Eq. (3). The protocol e is used to get a quick estimate of the efficiency of the method, although more detailed and complicated way could be used. The timing of each method requires the determination of the computational cost of each independent sample, which is defined by Eq. (5). The protocol f emphasizes two important points in using the stiff spring approximation. The first one is that the force constant should be sufficiently large to restrain the value of CV close enough to the desired value. The width of the distribution of CV is given by Eq. (15) and shown in Fig. S1 and S2. The second one emphasizes the need for a small ITS such as 1 fs, which provides a deeper understanding of the energetics of the system in the biased simulation and the stability of numerical integration of the equation of motion. Some quick estimates of acceptable ITS are provided by Eq. (20) and Eq. (25), the numerical results of which are plotted in Fig. 2. Therefore, the reason for using a time step as small as 1 fs is not the motion of hydrogen atoms, but the large force constant of harmonic potentials used in the stiff spring limit.

In a word, the above protocols provide a general nonequilibrium framework for the indirect QM free energy simulation, where the convergence can always be reached with a moderate sample size with a

fraction of the computational cost of the direct free energy simulation at QM level. Adjusting the framework could lead to some speedup under particular circumstances, e.g. using unidirectional pulling for the correction term. However, if someone wants a fail-safe method, which provides at least an order of magnitude speedup, the multi-dimensional nonequilibrium pulling framework is an option.

III. Computational Details

System preparation. Our test structural observable is one periodic CV of the backbone dihedral shown in Fig. 1b. This ACE-NME system in its solvated state is tested in our previous work and serves as a nice case for illustrative calculations. The system is constructed by caps in biomolecules, and is more biologically relevant than the C_2H_6 system. The SQM Hamiltonians of AM1,⁶² PM6⁶³ and MNDO⁶⁴ are selected as the low-level Hamiltonians, while HF,⁶⁵⁻⁶⁷ MP2,⁶⁸⁻⁷¹ B3LYP,⁷²⁻⁷⁴ and ω B97X-D⁷⁵ are employed for the high-level Hamiltonians. The basis set used in ab initio calculations is 6-31G*.

Free Energy Simulation. The sampling in the configurational space under all Hamiltonians is performed as follows. The dihedral windows or segments are equally spaced from 0° to 360° with 2° increments. The large force constant applied to achieve the stiff spring limit is $2000 \text{ kcal/mol}\cdot\text{rad}^2$, in order to define the initial state precisely. In the initial configurational sampling procedure, in each intermediate, we perform 1000 cycles energy-minimization with steepest descent algorithm, after which a 100 ps heating to 300 K is performed. After sufficient equilibration, the initial configurational sampling is performed with the sampling interval of 0.05 ps, which has been tested to be sufficient to decorrelate the successive samples in our previous work.^{27, 85} Bidirectional nonequilibrium pulling are then initiated from these uncorrelated configurations. For convergence check, we tested a series of pulling speeds, including 0.1 ps per 2° segment, 0.2 ps/segment, 0.5 ps/segment and 1 ps/segment.

As for the free energy calculation along the alchemical order parameter, the initial configurations are obtained from the initial configurational sampling procedure along the configurational CV mentioned above. In the alchemical free energy simulation, we tested a series of magnitudes for each perturbation of the alchemical control parameter, including $\Delta\lambda = 1$ (single-step perturbation of BAR), $\Delta\lambda = 0.5$, $\Delta\lambda = 0.1$, and $\Delta\lambda = 0.02$. The relaxation time between successive perturbations is 1 ITS.

The force constant used to achieve the stiff spring limit in the configurational space is set to $2000 \text{ kcal}/(\text{mol}\cdot\text{rad}^2)$, according to the above discussion about the width of CV distribution. As for the corresponding selection of ITS, as has been discussed above, 1 fs or 0.5 fs should be used when achieving the stiff spring limit. As this ITS is already smaller than the recommended 2 fs ITS when applying SHAKE⁸⁷

for bond-length constraints, we do not apply the bond-length constraints. We employ the time step 0.5 fs in our pulling simulation, although 1 fs has been tested to be also applicable in our previous work.^{27, 84-85, 89} Therefore, the pulling time along the alchemical CV is thus 0 fs, 1 fs, 5 fs, and 25 fs in each direction. Langevin dynamics⁹³ with the collision frequency of 5 ps⁻¹ are implemented for temperature regulation. As the simulation is performed in vacuo, there is no cutoff applied in our simulations. We use the AMBER⁸³ suite for MD simulation and Gaussian 09⁹⁴ for ab initio QM calculations. All statistical analyses are obtained with homemade codes.

IV. Result and discussion

Direct free energy simulations at SQM and QM levels.

We firstly check the convergence behavior of the direct free energy simulations at SQM and QM levels. The pulling-speed dependence and the sample-size dependence of the free energy profiles are shown in Fig. 3. The pulling-speed dependence of the free energy profiles under QM Hamiltonians not presented in Fig. 3 are provided in Fig. S3. Following our previous work,⁸⁵ we again use the Savitzky-Golay filter to increase the signal-to-noise ratio and smooth the PMF. From Fig. 3a, 3c, 3e, 3g and Fig. S3, we know that at all levels of theory, the pulling time for each 2° segment could not be smaller than 500 fs. Therefore, we assign this pulling speed as the pulling speed required for convergence. As for the sample-size dependence of the free energy profile shown in Fig. 3b, 3d, 3f and 3h, there are small (very small) variations in the first several iterations. With 20 samples, i.e. in the 4th iteration, the convergence could be well-achieved. Another proof of convergence can be obtained from the time-evolution of SD profile shown in Fig. S4. The non-linear and monotonically decreasing behavior of state-specified SDs suggests that the convergence under this sample size is well-achieved. Thus, the sample size required for convergence is assigned as 20. The resulting overall pulling time along the whole configurational pathway is 3.6 ns.

The correction term.

The correction terms obtained under different pulling speeds are shown in Fig. 4a and 4b. Here, we compare the unfitted data to provide a direct and convincing comparison of the data and avoid the contamination of the fitting procedure. The correction terms obtained from different pulling speeds (magnitude of perturbation per perturbation) are very similar, although those under slower pulling speeds suffer less from the noise. To avoid exaggerating the performance of the indirect scheme, we use the pulling speed of 1 fs, i.e. $\Delta\lambda = 0.5$, as the pulling speed required for convergence. Actually, the pulling time is only 0.5 fs, but we take 1 fs to avoid exaggerating the performance of our method.

The sample-size dependence of the correction term is shown in Fig. 4c and 4d. We can see that for both AM1 \leftrightarrow B3LYP correction and PM6 \leftrightarrow HF correction, the 3-sample, 10-sample and 50-sample PMFs do not differ much. Therefore, the convergence can be well achieved within 10 samples. To avoid exaggeration of the speedup of the indirect scheme, as has been pointed out in the protocol e in the method part, we use the equal sample size rule for everything. Namely, the sample size of 20 in the direct free energy simulation is also used as the sample size required for convergence in the SQM \leftrightarrow QM correction.

Table S1 provides the computational costs for the direct construction of free energy landscapes at SQM and QM levels as well as those in the SQM \leftrightarrow QM corrections, which may be a better and clearer way to compare the efficiency to computational scientists. We can see that the direct free energy simulation is very time-consuming, while the correction term is much faster than the direct simulations. The overall timing of the indirect method is thus much faster than the direct one.

Indirect vs direct.

After obtaining converged estimates from the direct and indirect free energy simulations, we then compare these free energy profiles. The comparison is provided in Fig. 5. We can see that for all target high-level Hamiltonians, the indirect results from the low-level corrections are able to reproduce the reference results. In Fig. 5a, the direct AM1, PM6 and MNDO PMFs obviously differ from the high-level one. By adding the SQM \leftrightarrow QM correction, the PMFs become much more similar to the direct B3LYP result. The free energy profiles at HF, MP2 and ω B97X-D levels are also reproduced by the indirect method, as is shown in Fig. 5b, 5c and 5d. The relative free energies at important points on the free energy profiles are summarized in Table S2, which enables a quantitative comparison. We can see that the indirect results indeed agree with the direct ones.

Compared with the previous case of performing the indirect free energy calculation at SQM level by combining the direct MM free energy simulation and the MM \leftrightarrow SQM correction in our previous work,⁸⁵ the agreement between the QM PMF and the SQM one is much better. The reason should be that the similarity between SQM and QM Hamiltonians is higher than the similarity between MM and SQM Hamiltonians, which, according to others' work, is mainly due to the differences in the stiff degree of freedoms.^{22, 95-99} The peaks in the SD profile under the MM Hamiltonian observed in our previous work could also be a sign of this.^{27, 85}

After checking the consistency of the free energy profiles from the direct and indirect methods, we then turn to compare the efficiencies of different methods. As the computational costs under different SQM Hamiltonians are extremely similar, we use their mean to estimate the general computational cost at SQM

level. The comparisons between computational costs of different schemes are given in Table 2 and Table S3, where the computational costs are scaled by the ratio of the computational cost under the QM Hamiltonian to that under the SQM Hamiltonian to estimate the equivalent simulation time under SQM Hamiltonians. From this comparison, we know that the speedup of the indirect scheme is generally about 20 fold. This phenomenon is drastically different from the previous case of performing the indirect free energy calculation at SQM level by using MM simulations and MM \leftrightarrow SQM corrections, where the computational speedup for this small toy system is at most 2 fold.⁸⁵ The reason is that the computational cost at ab initio QM level is significantly higher than the SQM one, while the computational costs under SQM and MM Hamiltonians are similar. As a result, the computational cost at SQM level is almost negligible, and the speedup of the indirect scheme is mainly determined by the ratio of the computational cost in the direct QM free energy simulation to that in the SQM \leftrightarrow QM correction, namely

$$\text{speedup} \approx \frac{2n_{\text{direct QM}} \phi_{\text{direct QM}}}{2n_{\text{SQM}\leftrightarrow\text{QM correction}} \phi_{\text{SQM}\leftrightarrow\text{QM correction}}} = \frac{\phi_{\text{NEW,direct QM}} + \phi_{\text{eq,direct QM}}}{\phi_{\text{NEW,SQM}\leftrightarrow\text{QM correction}} + \phi_{\text{eq,SQM}\leftrightarrow\text{QM correction}}} \quad (30)$$

. The factor of 2 in the above equation is due to the bidirectionality of the pulling, and the equal sample size rule for all nonequilibrium transformations defined in our protocol e leads to $n_{\text{direct QM}} = n_{\text{SQM}\leftrightarrow\text{QM correction}}$. If we overestimate the sample size required for convergence in the direct free energy simulation by using $n_{\text{direct QM}} > n_{\text{SQM}\leftrightarrow\text{QM correction}}$, the speedup of the indirect scheme is overestimated significantly. Such an overestimation often happens when one uses a large number of samples to estimate the sample size required for convergence in the direct free energy simulation, but uses a subset of the whole dataset to calculate the SQM \leftrightarrow QM correction. As long as the samples are independent, it is not proper to differentiate different samples. An illustration of such an overestimation is provided in Table 3, where the sample sizes in the direct free energy simulations at SQM and QM level are overestimated to 100 per segment and the sample sizes required for convergence in the SQM \leftrightarrow QM are underestimated to 10 per segment. The resulting overall sampling time in the direct free energy simulation under each Hamiltonian is 18.9 ns, rather than the 3.78 ns in Table 2. The speedup of the indirect scheme, obviously, is significantly overestimated. The 200-fold speedup in Table 3 is quite appealing, but is unreal and unfair due to the inappropriate processing of the timing data. However, by applying the protocol e, we avoid such an overestimation and provide a fair estimate of the acceleration. Namely, all independent samples should be used or at least be treated fairly to calculate this speedup.

As the statistical inefficiency in the direct QM free energy calculation and that in the SQM \leftrightarrow QM correction are the same, the equation can be reduced to

$$\text{speedup} \approx \frac{\phi_{\text{NEW,direct QM}} + \phi_{\text{eq}}}{\phi_{\text{NEW,SQM}\leftrightarrow\text{QM correction}} + \phi_{\text{eq}}} \quad (31)$$

. Further, as the statistical inefficiency in equilibrium states is a constant for a given, the only variables in the above equation are the pulling speeds of $\phi_{\text{NEW,direct QM}}$ and $\phi_{\text{NEW,SQM}\leftrightarrow\text{QM correction}}$. The pulling speed to reach convergence is positively correlated to the similarity of different states and is negatively correlated to the magnitude of perturbation during the pulling process. Thus, in the configurational space, a high similarity between neighboring configurational states and a smaller perturbation in jumping between neighboring configurational states would lead to a less efficient indirect scheme, while a low similarity between neighboring configurational states and a large perturbation in pulling between them would make the indirect scheme efficient. As for the alchemical space, if the low-level Hamiltonian is sufficiently similar to the high-level one, the indirect scheme would be extremely efficient.

Comparison between SQM Hamiltonians.

In Fig. 5, although all indirect results are very similar to the direct ones, the degrees of agreement differ. An obvious phenomenon is that the indirect results from PM6 reproduce the direct results very well, while the other agree with the direct results but the deviations are relatively large. This provides some information about the selection of the low-level Hamiltonian. Among SQM Hamiltonians tested here, PM6 would be the best choice for the low-level Hamiltonians in indirect free energy simulation for ab initio QM results. AM1 and MNDO are relatively worse and may not be useful.

A worth noting point is that the motivation of using AM1 as the low-level Hamiltonian, to some extent, is its ability to capture the basic quantum effect. The widely used charge model derived from AM1 is AM1-BCC,¹⁰⁰⁻¹⁰¹ where BCC is the acronym of bond charge correction. A linear correction added to the AM1 charge is good enough to reproduce the electrostatic potential (ESP) around the molecule at HF/6-31G* level, which inspires us that the AM1 free energy landscape should be able to reproduce the HF PMF by adding a correction term. In the current case, the correction method does work. However, the agreement is a bit worse than the indirect result obtained from PM6, which indicates that PM6 is more similar to the high-level Hamiltonians. Therefore, following this idea, perhaps deriving the PM6-BCC charge could provide a better charge model than the AM1-BCC one.

V. Conclusion

Ab initio QM calculations are time-consuming. Converged estimates of free energy landscapes from free energy simulations are computationally demanding at QM level. The indirect scheme for the construction of free energy landscapes at QM level constructs an alternative transformation pathway. By combining the direct PMF for the transformation under the low-level Hamiltonian and the low-level-to-high-level correction, the free energy landscapes at QM level can be obtained. In this work, we extend our previously proposed multi-dimensional nonequilibrium framework for indirect QM/MM free energy simulation to the calculation at ab initio QM level, where the direct SQM free energy simulation is performed and the correction term accounts for the SQM \leftrightarrow QM perturbation. The agreement between the indirect results and the direct ones is very good. Further, we emphasize the importance of not exaggerating the performance of the indirect scheme again. The speedup of the indirect scheme could be overestimated by an order of magnitude by inappropriate calculations of the sample size required for convergence in different free energy simulation schemes.

We tested a series of SQM Hamiltonians (AM1, PM6 and MNDO) and QM Hamiltonians (HF, MP2, B3LYP and ω B97X-D), which could provide some guidance for the selection of the low-level SQM methods to enhance the convergence of the indirect method for other systems in the future. It is observed that the indirect results obtained from PM6 show the best agreement to the high-level QM PMFs, while AM1 and MNDO perform less satisfactorily. Therefore, in selecting the low-level theory in performing the indirect free energy simulations, PM6 would be the best choice. Further, the relative performance of the different SQM Hamiltonians could provide some guidance for the fitting of atomic charges. As is widely known, the AM1-BCC charge model provides a cheaper yet accurate alternative to the restrained ESP (RESP) fitting, and is able to reproduce the ESP around the molecule accurately. However, in our work, PM6 is observed to have a higher similarity to the high-level Hamiltonians than AM1. Therefore, perhaps a new charge model can be fitted based on the PM6 Hamiltonian, which can be called as PM6-BCC. This charge could, in principle, provide a more accurate charge model compared with the existing AM1-BCC one.

The key features of the nonequilibrium framework for indirect QM free energy simulations are the nonequilibrium pulling and bidirectional reweighting. The first one ensures the convergence can be reached with not-that-huge sample size and the sample size hysteresis is no longer a problem. The second one makes the statistically optimal use of the samples and also helps with the sample size hysteresis. The theoretically rigorous definition of the computational cost of each independent sample makes the analytical uncertainty meaningful. In some circumstances, altering the framework such as using unidirectional pulling coupled to

JI may lead to some further acceleration. However, if a simple-to-apply and fail-safe method is expected, use the current multi-dimensional pulling one. Further, compared to equilibrium dynamics such as umbrella sampling, the multi-stage nature of the nonequilibrium pulling framework makes each pulling segment very short, which can be finished within a reasonable time with only a single core. As a result, the nonequilibrium pulling simulation is highly parallel, which maximizes the efficiency of the computing power. No openMP- or MPI-related issue exists to degrade the efficiency of the computation.

Another central result of the current work focuses on the estimation of the force constant used in umbrella-sampling-like simulations and the selection of corresponding time step in the integration of the equation of motions. For small force constants used in umbrella sampling simulations, the ITS could be set to 2 fs if the bond-length constraints (e.g. SHAKE) are used. However, if a large force constant is used to achieve the stiff spring limit, a small time step (e.g. 1 fs or 0.5 fs) must be used to ensure numerical stability to integrate the equation of motions, even with the SHAKE option on.

Acknowledgement

This work was supported China Scholarship Council. Computer access to the CLAIX cluster of RWTH Aachen University and clusters of Forschungszentrum Juelich is gratefully acknowledged. We are grateful for many valuable and insightful comments from Prof. Dr. Piero Procacci (University of Florence) and the anonymous reviewers.

Conflicts of interest

There are no conflicts of interest to declare.

Supporting information

The scanned width of CV distribution under different force constants and temperatures for dihedral and distance CVs, the time-evolution of dimensionless SD profiles in nonequilibrium stratification and timing comparison between the direct and indirect schemes for different combinations of low-level and high-level Hamiltonians are given.

References

1. Echeverria, I.; Amzel, L. M., Helix propensities calculations for amino acids in alanine based peptides using Jarzynski's equality. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78* (5), 1302-1310.
2. Lee, T. S.; Radak, B. K.; Huang, M.; Wong, K. Y.; York, D. M., Roadmaps through free energy landscapes calculated using the multi-dimensional vFEP approach. *J Chem Theory Comput* **2014**, *10* (1), 24-34.
3. Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chem. Phys. Lett.* **2017**, *684*, 239-244.
4. Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V., Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (11), E2136.
5. Mezei, M., Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias. *J. Comput. Phys.* **1987**, *68* (1), 237-248.
6. Hooft, R. W.; van Eijck, B. P.; Kroon, J., An Adaptive Umbrella Sampling Procedure in Conformational Analysis using Molecular Dynamics and Its Application to Glycol. *J. Chem. Phys.* **1992**, *97* (9), 6690-6694.
7. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1* (6), 932-942.
8. Hub, J. S.; Groot, B. L. D.; Spoel, D. V. D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Phys.* **2015**, *6* (9), 3713-3720.
9. Wood, R. H.; Muhlbauer, W. C. F.; Thompson, P. T., Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis. *J. Phys. Chem.* **1991**, *95* (17), 6670-6675.
10. Gore, J.; Ritort, F.; Bustamante, C., Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci U S A* **2003**, *100* (22), 12564-9.
11. Zuckerman, D. M.; Woolf, T. B., Theory of a systematic computational error in free energy differences. *Phys Rev Lett* **2002**, *89* (18), 180602.
12. Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S., Equilibrium Free Energies from Nonequilibrium Measurements using Maximum-likelihood Methods. *Phys. Rev. Lett.* **2003**, *91* (14), 140601.
13. Bennett, C. H., Efficient Estimation of Free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22* (2), 245-268.
14. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105.
15. Lee, T. S.; Radak, B. K.; Pabis, A.; York, D. M., A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J Chem Theory Comput* **2013**, *9* (1), 153-164.
16. Jarzynski, C., A Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78* (14), 2690-2693.
17. Mallick, K.; Moshe, M.; Orland, H., Supersymmetry and Nonequilibrium Work Relations. *arXiv preprint arXiv:0711.2059* **2008**.
18. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Criteria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.
19. Ballard, A. J.; Jarzynski, C., Replica exchange with nonequilibrium switches: enhancing equilibrium sampling by increasing replica overlap. *J. Chem. Phys.* **2012**, *136* (19), 194101.
20. Vaikuntanathan, S.; Jarzynski, C., Escorted free energy simulations: improving convergence by reducing dissipation. *Phys. Rev. Lett.* **2008**, *100* (19), 190601.
21. Dickson, A.; Dinner, A. R., Enhanced Sampling of Nonequilibrium Steady States. *Annual Review of Physical Chemistry* **2010**, *61* (1), 441-459.
22. Hudson, P. S.; Woodcock, H. L.; Boresch, S., Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems. *J Phys Chem Lett* **2015**, *6* (23), 4850-6.
23. Procacci, P.; Marsili, S., Energy dissipation asymmetry in the non equilibrium folding/unfolding of the single molecule alanine decapeptide. *Chem. Phys.* **2010**, *375* (1), 8-15.

24. Pham, T. T.; Shirts, M. R., Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *J. Chem. Phys.* **2011**, *135* (3), 034114.
25. Sun, Z. X.; Wang, X. H.; Zhang, J. Z. H., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation. *Phys. Chem. Chem. Phys.* **2017**, *19* (23), 15005-15020.
26. Wang, X.; Tu, X.; Zhang, J. Z. H.; Sun, Z., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation: The Nonequilibrium Stratification. *Phys. Chem. Chem. Phys.* **2018**, *20* (3), 2009-2021.
27. Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019**, *9999*, 1-20.
28. Bruckner, S.; Boresch, S., Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32* (7), 1320–1333.
29. Resat, H.; Mezei, M., Studies on free energy calculations. I. Thermodynamic integration using a polynomial path. *J. Chem. Phys.* **1993**, *99* (8), 6052-6061.
30. Resat, H.; Mezei, M., Studies on free energy calculations. II. A theoretical approach to molecular solvation. *J. Chem. Phys.* **1994**, *101* (7), 6126-6140.
31. Paliwal, H.; Shirts, M. R., A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *J Chem Theory Comput* **2011**, *7* (12), 4115-34.
32. Fenwick, M. K.; Escobedo, F. A., On the use of Bennett's acceptance ratio method in multi-canonical-type simulations. *J Chem Phys* **2004**, *120* (7), 3066-74.
33. Sun, Z.; Zhu, T.; Wang, X.; Mei, Y.; Zhang, J. Z., Optimization of convergence criteria for fragmentation methods. *Chem. Phys. Lett.* **2017**, *687*, 163-170.
34. Liu, W.; Sakane, S.; And, R. H. W.; Doren, D. J., The Hydration Free Energy of Aqueous Na⁺ and Cl⁻ at High Temperatures Predicted by ab Initio/Classical Free Energy Perturbation: 973 K with 0.535 g/cm³ and 573 K with 0.725 g/cm³. *J.phys.chem.a* **2002**, *106* (7), 1409-1418.
35. Olsson, M. A.; Söderhjelm, P.; Ryde, U., Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J. Comput. Chem.* **2016**, *37* (17), 1589-1600.
36. Raghavachari, K.; Saha, A., Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev.* **2015**, *115* (12), 5643-5677.
37. Collins, M. A.; Bettens, R. P., Energy-based molecular fragmentation methods. *Chem. Rev.* **2015**, *115* (12), 5607-5642.
38. Sahu, N.; Gadre, S. R., Molecular tailoring approach: a route for ab initio treatment of large clusters. *Accounts of chemical research* **2014**, *47* (9), 2739-2747.
39. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M., Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313* (3-4), 701-706.
40. Gao, J.; Luque, F. J.; Orozco, M., Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solution. *J. Chem. Phys.* **1993**, *98* (4), 2975-2982.
41. †, V. L.; Warshel, A., Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J. Comput. Chem.* **1992**, *13* (2), 199–213.
42. Wesolowski, T.; Warshel, A., Ab Initio Free Energy Perturbation Calculations of Solvation Free Energy Using the Frozen Density Functional Approach. *J. Phys. Chem.* **1994**, *98* (20), 5183-5187.
43. Gao, J.; Xia, X., A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258* (5082), 631-5.
44. Zheng, Y. J.; Merz, K. M., Mechanism of the human carbonic anhydrase II-catalyzed hydration of carbon dioxide. *J. Am. Chem. Soc.* **1992**, *114* (26), 10498-10507.
45. Lameira, J. S.; Kupchenko, I.; Warshel, A., Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J. Phys. Chem. B* **2016**, *120* (9), 2155.
46. Plotnikov, N. V.; Warshel, A., Exploring, refining, and validating the paradynamics QM/MM sampling. *J. Phys. Chem. B* **2012**, *116* (34), 10342-10356.
47. Plotnikov, N.; Kamerlin, S. C. L.; Warshel, A., ParaDynamics: An Effective and Reliable Model for Ab Initio QM/MM Free

Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115* (24), 7950-62.

48. Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A., Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *The Journal of Physical Chemistry B* **1998**, *102* (12), 2293-2301.
49. Polyak, I.; Benighaus, T.; Boulanger, E.; Thiel, W., Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation. *J. Chem. Phys.* **2013**, *139* (6), 578.
50. König, G.; Boresch, S., Non-Boltzmann sampling and Bennett's acceptance ratio method: How to profit from bending the rules. *J. Comput. Chem.* **2011**, *32* (6), 1082-90.
51. Heimdal, J.; Rydberg, P.; Ryde, U., Protonation of the proximal histidine ligand in heme peroxidases. *J. Phys. Chem. B* **2008**, *112* (8), 2501-10.
52. Mikulskis, P.; Cioloboc, D.; Andrejić, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U., Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 375-400.
53. Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O.; Essex, J. W.; Skylaris, C. K., Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies. *J. Phys. Chem. B* **2013**, *117* (32), 9478-85.
54. Genheden, S.; Ryde, U.; Söderhjelm, P., Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J. Comput. Chem.* **2015**, 2114-2124.
55. Genheden, S.; Martinez, A. I. C.; Criddle, M. P.; Essex, J. W., Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 187-200.
56. Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C. K., Free Energies of Binding from Large-Scale First-Principles Quantum Mechanical Calculations: Application to Ligand Hydration Energies. *J. Phys. Chem. B* **2013**, *117* (32), 9478-85.
57. Woods, C. J.; Manby, F. R.; Mulholland, A. J., An efficient method for the calculation of quantum mechanics/molecular mechanics free energies. *J. Chem. Phys.* **2008**, *128* (1), 152-159.
58. Caveayland, C.; Skylaris, C. K.; Essex, J. W., Direct Validation of the Single Step Classical to Quantum Free Energy Perturbation. *J. Phys. Chem. B* **2014**, *119* (3), 1017-25.
59. König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L., 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.* **2014**, *10* (4), 1406-1419.
60. Rod, T. H.; Ryde, U., Quantum mechanical free energy barrier for an enzymatic reaction. *Phys. Rev. Lett.* **2005**, *94* (13).
61. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L., Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2015**, *29* (5), 397-411.
62. Dewar, M. J. S.; Zebisch, E. G.; Healy, E. F.; Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902-3909.
63. Stewart, J. J., Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13* (12), 1173-1213.
64. Dewar, M. J. S.; Thiel, W., Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899-4907.
65. Roothaan, C. C. J., New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23* (23), 69-89.
66. Pople, J. A.; Nesbet, R. K., Self - Consistent Orbitals for Radicals. *J. Chem. Phys.* **1954**, *22* (3), 571-572.
67. Mcweeny, R.; Dierksen, G., Self - Consistent Perturbation Theory. II. Extension to Open Shells. *J. Chem. Phys.* **1968**, *49* (11), 4852-4856.
68. Head-Gordon, M.; Head-Gordon, T., Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.* **1994**, *220* (1-2), 122-128.
69. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., A direct MP2 gradient method. *Chem. Phys. Lett.* **1990**, *166* (3), 275-280.
70. Head-Gordon, M.; Pople, J. A.; Frisch, M. J., MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153* (6),

503-506.

71. Frisch, M. J.; Head-Gordon, M.; Pople, J. A., Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.* **1990**, *166* (3), 281-289.
72. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98* (45), 11623-11627.
73. Becke, A. D., Density - functional thermochemistry. IV. A new dynamical correlation functional and implications for exact - exchange mixing. *J. Chem. Phys.* **1996**, *104* (3), 1040-1046.
74. Hertwig, R. H.; Koch, W., On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chem. Phys. Lett.* **1997**, *268* (5), 345-351.
75. Chai, J. D.; Head-Gordon, M., Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128* (8), 57-63.
76. Balsera; Stepaniants; Izrailev; Oono; Schulten, Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.* **1997**, *73* (3), 1281.
77. Marsili, S.; Procacci, P., Free energy reconstruction in bidirectional force spectroscopy experiments: The effect of the device stiffness. *The Journal of Physical Chemistry B* **2010**, *114* (7), 2509-2516.
78. Hummer, G.; Szabo, A., From the Cover: Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Science* **2001**, *98* (7), 3658-3661.
79. Hummer, G.; Szabo, A., Free Energy Reconstruction from Nonequilibrium Single-molecule Pulling Experiments. *Proceedings of the National Academy of Sciences* **2001**, *98* (7), 3658-3661.
80. Hummer, G.; Szabo, A., Free Energy Surfaces from Single-molecule Force Spectroscopy. *Cheminform* **2005**, *36* (39), 504-513.
81. Paramore, S.; Ayton, G. S.; Voth, G. A., Extending the Fluctuation Theorem to Describe Reaction Coordinates. *J. Chem. Phys.* **2007**, *126* (5), 051102.
82. Sun, Z.; Wang, X.; Song, J., Extensive Assessment of Various Computational Methods for Aspartate's pKa Shift. *J. Chem. Inf. Model.* **2017**, *57* (7), 1621-1639.
83. Case, D. A.; Cheatham, T. E.; Tom, D.; Holger, G.; Luo, R.; Merz, K. M.; Alexey, O.; Carlos, S.; Bing, W.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668-1688.
84. Wang, X.; Sun, Z., Determination of Base Flipping Free Energy Landscapes from Nonequilibrium Stratification. *J. Chem. Inf. Model.* **2019**.
85. Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019**, *21* (12), 6672-6688
86. Tuckerman, M. E.; Berne, B. J.; Martyna, G. J., Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *J. Chem. Phys.* **1991**, *94* (10), 6811-6815.
87. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of The Cartesian Equations of Motion of A System with Constraints: Molecular Dynamics of n -alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341.
88. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of The SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952-962.
89. Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019**.
90. Marchi, M.; Procacci, P., Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the NPT ensemble. *J. Chem. Phys.* **1998**, *109* (13), 5194-5202.
91. Procacci, P.; Marchi, M., Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm. *J. Chem. Phys.* **1996**, *104* (8), 3003-3012.
92. Procacci, P.; Berne, B., Computer simulation of solid C60 using multiple time - step algorithms. *J. Chem. Phys.* **1994**, *101* (3), 2421-2431.
93. Pastor, R. W.; Brooks, B. R.; Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **1988**, *65* (6), 1409-1419.
94. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson,

- G., GAUSSIAN09, Gaussian, Inc., Wallingford, CT, USA,(2009). *Google Scholar* **2016**.
95. Heimdal, J.; Ryde, U., Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys Chem Chem Phys* **2012**, *14* (36), 12592-604.
96. Konig, G.; Brooks, B. R., Correcting for the free energy costs of bond or angle constraints in molecular dynamics simulations. *Biochim Biophys Acta* **2015**, *1850* (5), 932-43.
97. Hudson, P. S.; White, J. K.; Kearns, F. L.; Hodoscek, M.; Boresch, S.; Lee Woodcock, H., Efficiently computing pathway free energies: New approaches based on chain-of-replica and Non-Boltzmann Bennett reweighting schemes. *Biochim Biophys Acta* **2015**, *1850* (5), 944-53.
98. Genheden, S.; Cabedo Martinez, A. I.; Criddle, M. P.; Essex, J. W., Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. *J Comput Aided Mol Des* **2014**, *28* (3), 187-200.
99. Konig, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L., 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* **2014**, *10* (4), 1406-1419.
100. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high - quality atomic charges. AM1 - BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23* (16), 1623-1641.
101. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high - quality atomic charges. AM1 - BCC model: I. Method. *J. Comput. Chem.* **2000**, *21* (2), 132-146.

Table 1. The timing information (in ns/day) of SQM and QM calculations on a computing node with 384 GB memory and the CPU used is Intel(R) Xeon(R) Platinum 8160. A single core is used for this benchmark, which avoids the influence of parallelization-related issues in the implementation of software. The memory available on each computing node is very large. Therefore, the memory-related issue in QM calculations should also not influence the speed of calculation. For ab initio QM calculation, the basis set of 6-31G* is employed. We can see that different SQM simulations are of very similar computational costs, while the speed of ab initio QM calculations depends on the level of theory.

Level of Theory Terms	semi-empirical QM				ab initio QM		
	AM1	PM6	MNDO	HF	B3LYP	MP2	wB97XD
speed(ns/day)	46.5	46.44	46.78	0.013907	0.002881	0.006472	0.001898
speedup(SQM/QM)	1.00	1.00	0.99	3343.64	16140.23	7184.80	24499.47

Table 2. Efficiency comparison of direct and indirect free energy simulations at QM level. The total simulation time in direct scheme is given by $N_{\text{segments}} * N_{\text{traj}} * (\phi_{\text{NEW}} + \phi_{\text{eq}})$, while the total simulation time in the indirect scheme is the sum of $N_{\text{segments,SQM}} * N_{\text{traj,SQM}} * (\phi_{\text{NEW,SQM}} + \phi_{\text{eq,SQM}})$ at SQM level and $N_{\text{traj,SQM}\rightarrow\text{QM}} * (\phi_{\text{NEW,SQM}\rightarrow\text{QM}} + \phi_{\text{eq,SQM}}) + N_{\text{traj,QM}\rightarrow\text{SQM}} * (\phi_{\text{NEW,QM}\rightarrow\text{SQM}} + \phi_{\text{eq,QM}})$ in the SQM \leftrightarrow QM correction. N_{segments} is the number of segments and N_{traj} is the number of realizations per segment. The simulation time at QM level is scaled by the ratio of computational cost under QM Hamiltonian and that under SQM Hamiltonian in Table 1 to provide the effective simulation time at SQM level, enabling the direct comparison between computational costs from different levels of theory. The computational cost of SQM \rightarrow QM differs from that of QM \rightarrow SQM, as the initial configuration sampling procedures proceed under different Hamiltonians. We select the DFT level of B3LYP as an example to illustrate the timing comparison, while the others are included in the supporting information.

Terms Simulation	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ps) scaled to SQM Hamiltonian	Relative efficiency
direct SQM	0.05	0.5x2=1	180	20	3780.00	16140.23
SQM \rightarrow B3LYP	0.05	0.001	180	20	58284.82	-
B3LYP \rightarrow SQM	0.05	0.001	180	20	2963346.06	-
indirect B3LYP	-	-	-	-	3025410.89	20.17
direct B3LYP	0.05	0.5x2=1	180	20	61010065.95	1.00

Table 3. Efficiency comparison of direct and indirect free energy simulation at QM level when the sample size in the direct free energy simulation is overestimated to 100 per segment and that in the SQM \leftrightarrow QM correction is underestimated to 10 per segment. Total simulation time in direct scheme is given by $N_{\text{segments}} * N_{\text{traj}} * (\phi_{\text{NEW}} + \phi_{\text{eq}})$, while the total simulation time in the indirect scheme is the sum of $N_{\text{segments,SQM}} * N_{\text{traj,SQM}} * (\phi_{\text{NEW,SQM}} + \phi_{\text{eq,SQM}})$ at SQM level and $N_{\text{traj,SQM}\rightarrow\text{QM}} * (\phi_{\text{NEW,SQM}\rightarrow\text{QM}} + \phi_{\text{eq,SQM}}) + N_{\text{traj,QM}\rightarrow\text{SQM}} * (\phi_{\text{NEW,QM}\rightarrow\text{SQM}} + \phi_{\text{eq,QM}})$ in SQM \leftrightarrow QM correction. N_{segments} is the number of segments and N_{traj} is the number of realizations per segment. The simulation time at QM level is scaled by the ratio of computational cost under QM Hamiltonian and that under SQM Hamiltonian in Table 1 to be the effective simulation time at SQM level, enabling direct comparison between computational costs. The computational cost of SQM \rightarrow QM differs from QM \rightarrow SQM, as the initial configuration sampling procedures proceed under different Hamiltonians.

Terms Simulation	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ps) scaled to SQM Hamiltonian	Relative efficiency
direct SQM	0.05	0.5x2=1	180	100	18900.00	16140.23
SQM \rightarrow B3LYP	0.05	0.001	180	10	29142.41	-
B3LYP \rightarrow SQM	0.05	0.001	180	10	1481673.03	-
indirect B3LYP	-	-	-	-	1529715.44	199.42
direct B3LYP	0.05	0.5x2=1	180	100	305050329.75	1.00

Fig. 1. a) An illustration of multi-dimensional free energy calculation. As we define k_1 as the configurational CV, k_2 as the alchemical CV, $K_2 = 1$ as the SQM state and $K_2 = 2$ as the QM state, the definition of each state in the 2D free energy simulation is (k_1, k_2) . The direct SQM free energy simulation is the transformation from $(1, 2)$ to $(k_1, 2)$ and the indirect scheme is the combination of the SQM transformation $(1, 1) \leftrightarrow (k_1, 1)$ and SQM \leftrightarrow QM correction $(k_1, 1) \leftrightarrow (k_1, 2)$. In the indirect scheme, the goal is to get the free energy landscape at high-level Hamiltonian. Therefore, according to efficiency considerations in nonequilibrium simulation, only the transformations described with solid arrows are performed. b) The illustration of the dihedral flipping process. The definition of the reaction coordinate is the backbone dihedral in ACE-NME (C-C-N-C). (Red O atom, cyan C, white H, blue N atom.)

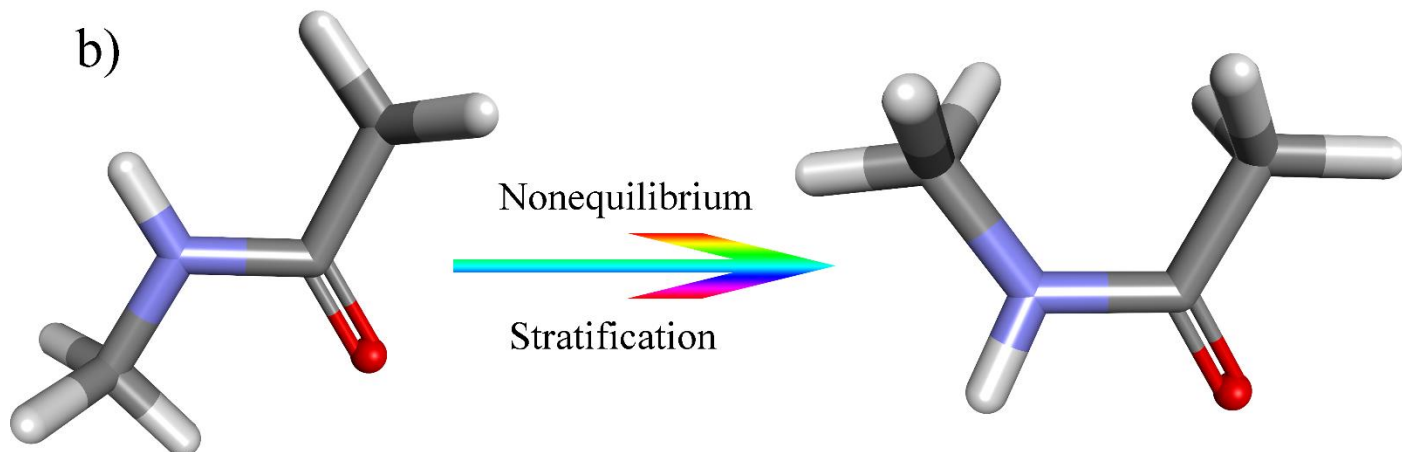
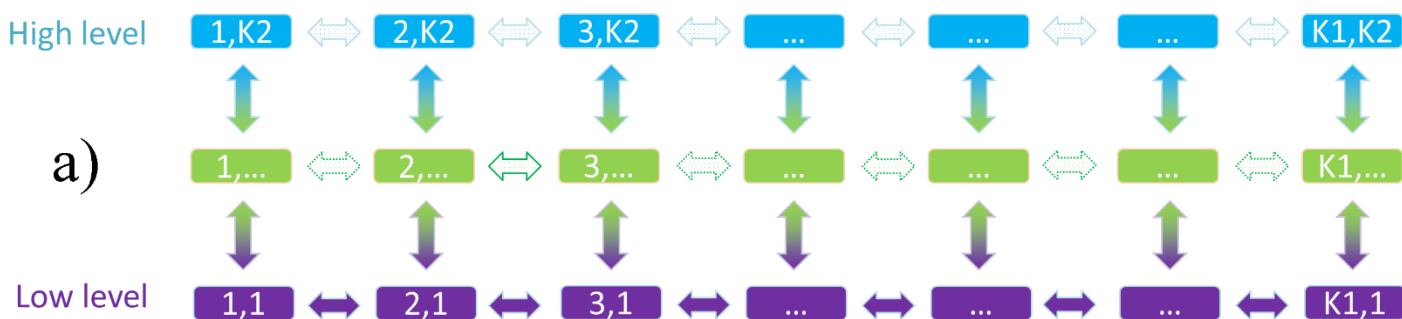


Fig. 2. a) Acceptable time step as a function of force constant in AMBER at 300 K in the dihedral case. The effective mass is set to 10 g/mol. Values between the two curves are usable for time steps in MD simulations. b) Acceptable time step as a function of force constant in AMBER at 300 K in the distance case. The mass is set to 10 g/mol. Values between the two curves are usable for time steps in MD simulations.

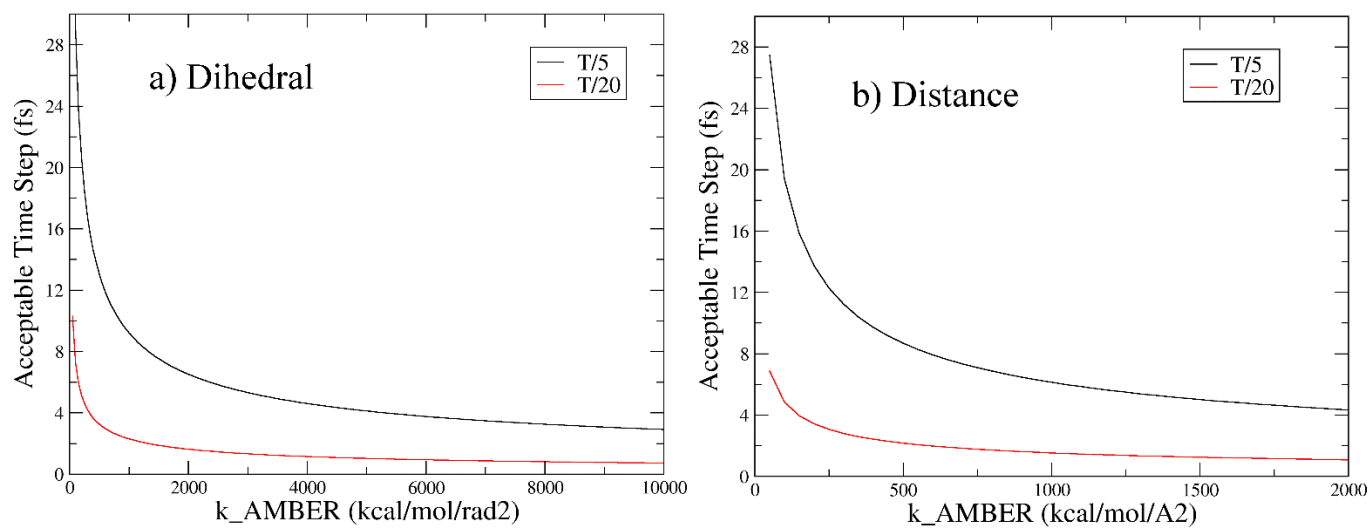
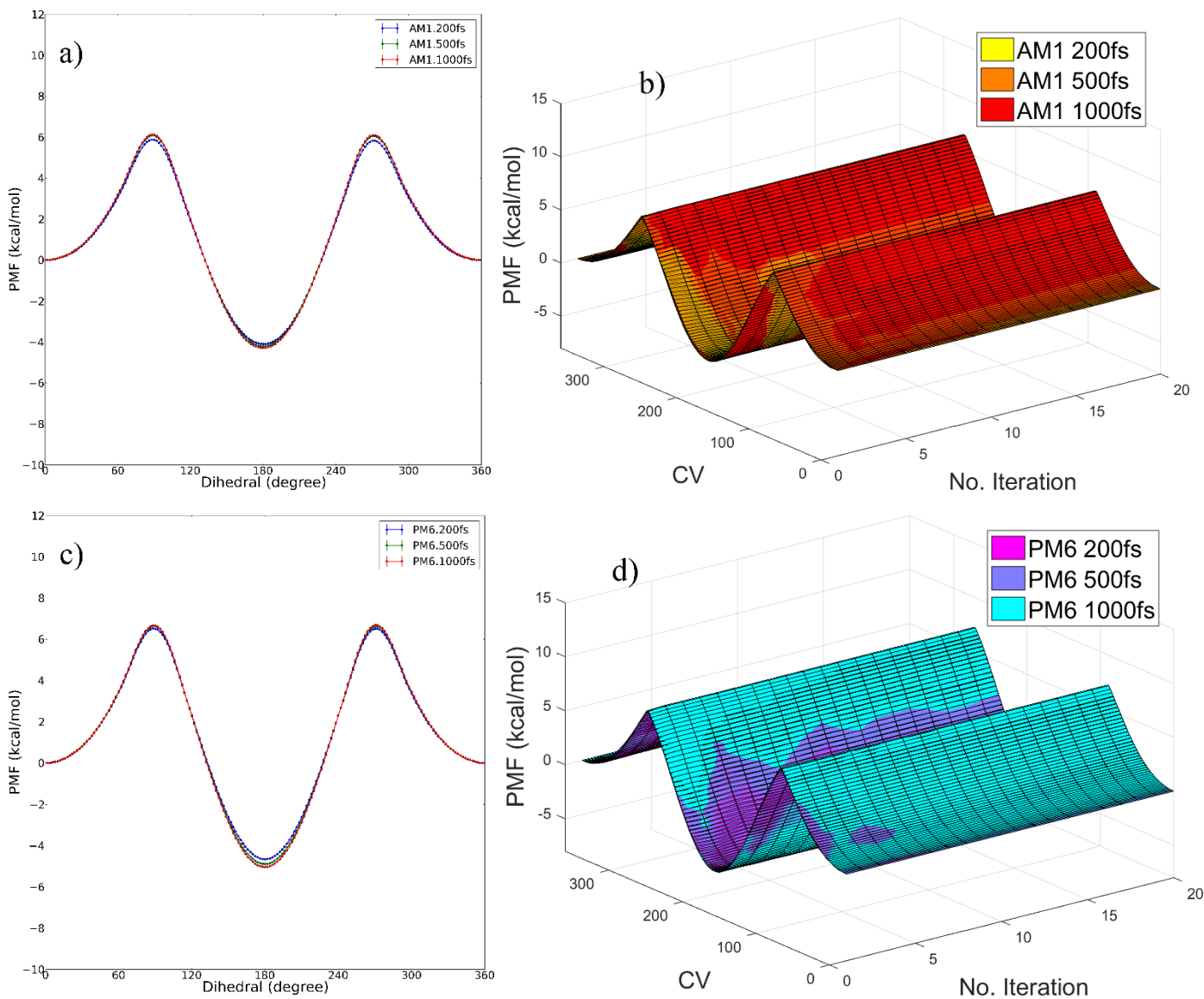


Fig. 3. Under different Hamiltonians, the convergence behavior of free energy profiles on the sample size and pulling speeds at a-b) AM1, c-d) PM6, e-f) MNDO, and g-h) B3LYP levels. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. The time in the legend represents the pulling time for each 2° segment. From the pulling-speed dependence, we know that 0.5 ps per segment is slow enough for convergence and thus the statistics under this pulling speed is used for discussion in the following parts of the paper.



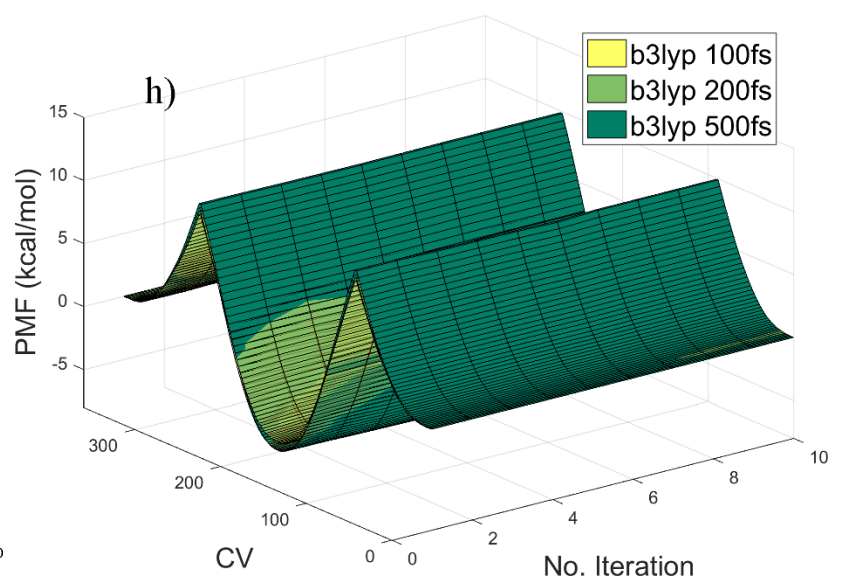
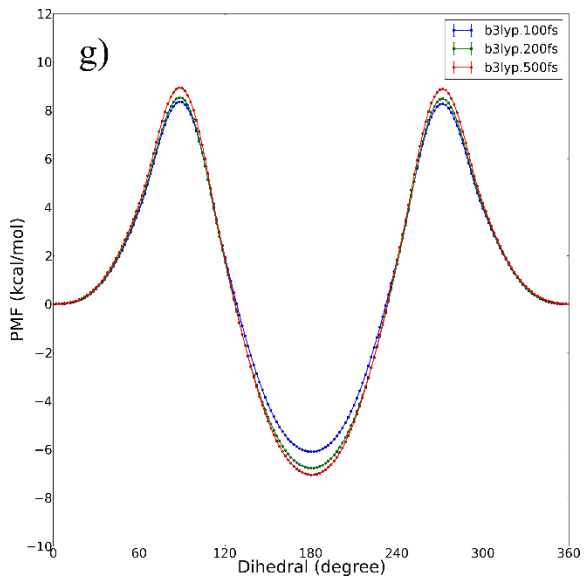
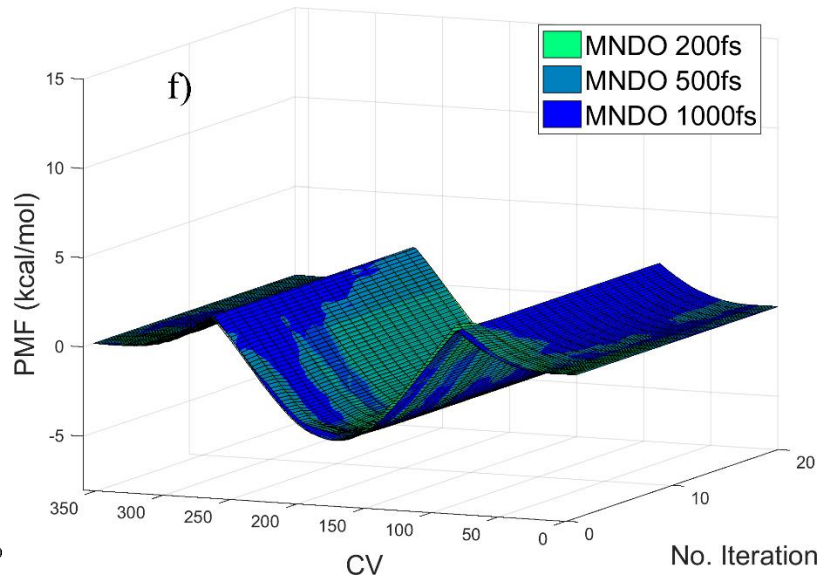
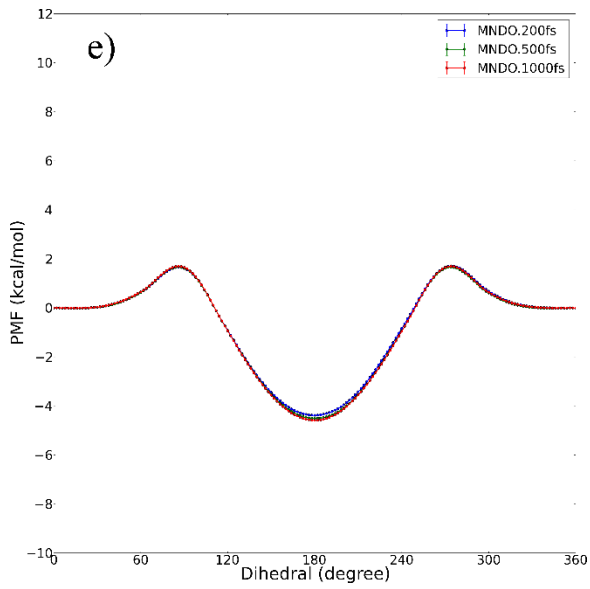


Fig. 4. Convergence of the SQM \leftrightarrow QM correction on the pulling speeds and sample size. a) and c) is for the AM1 \leftrightarrow B3LYP correction, while b) and d) are for the PM6 \leftrightarrow HF correction.

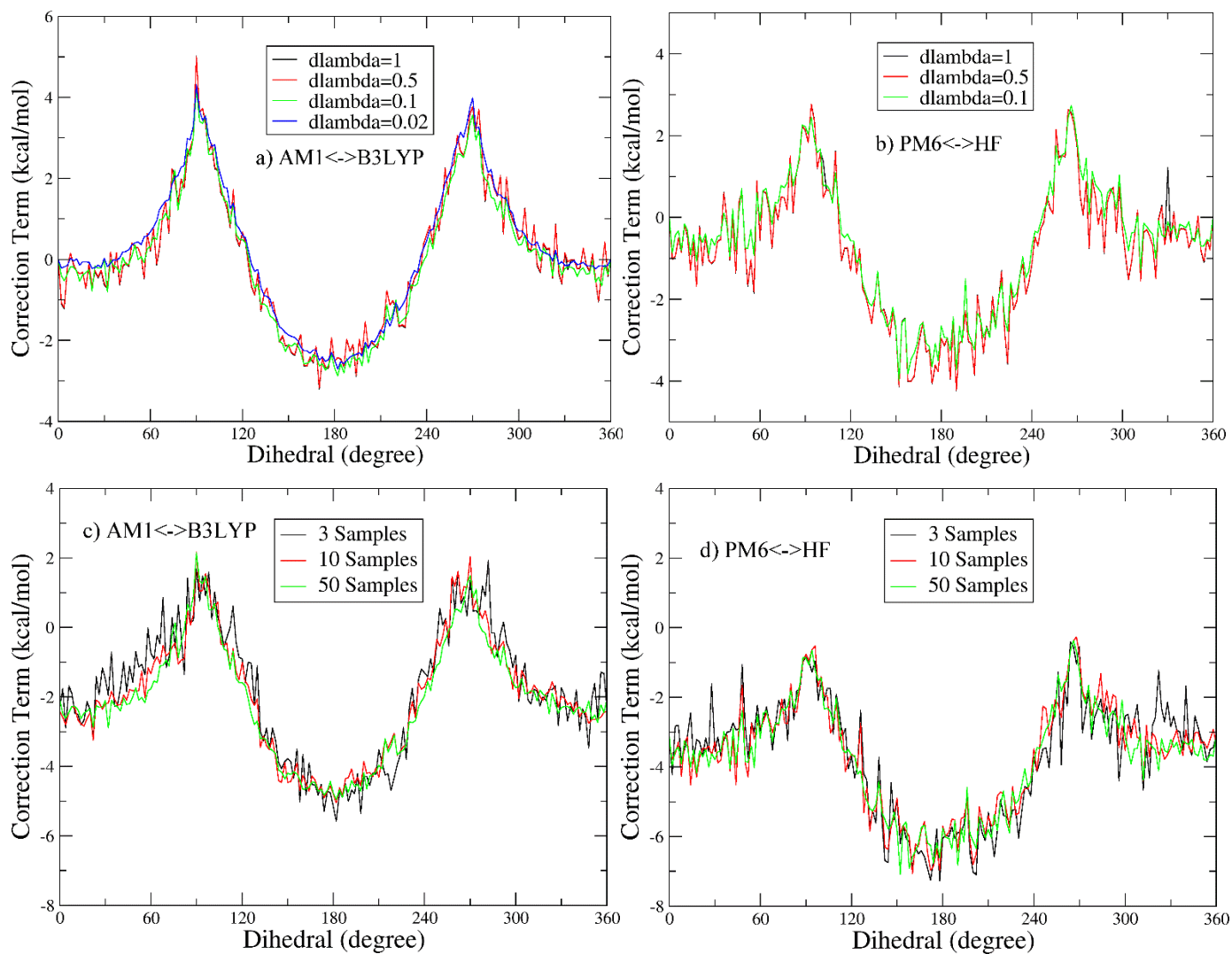
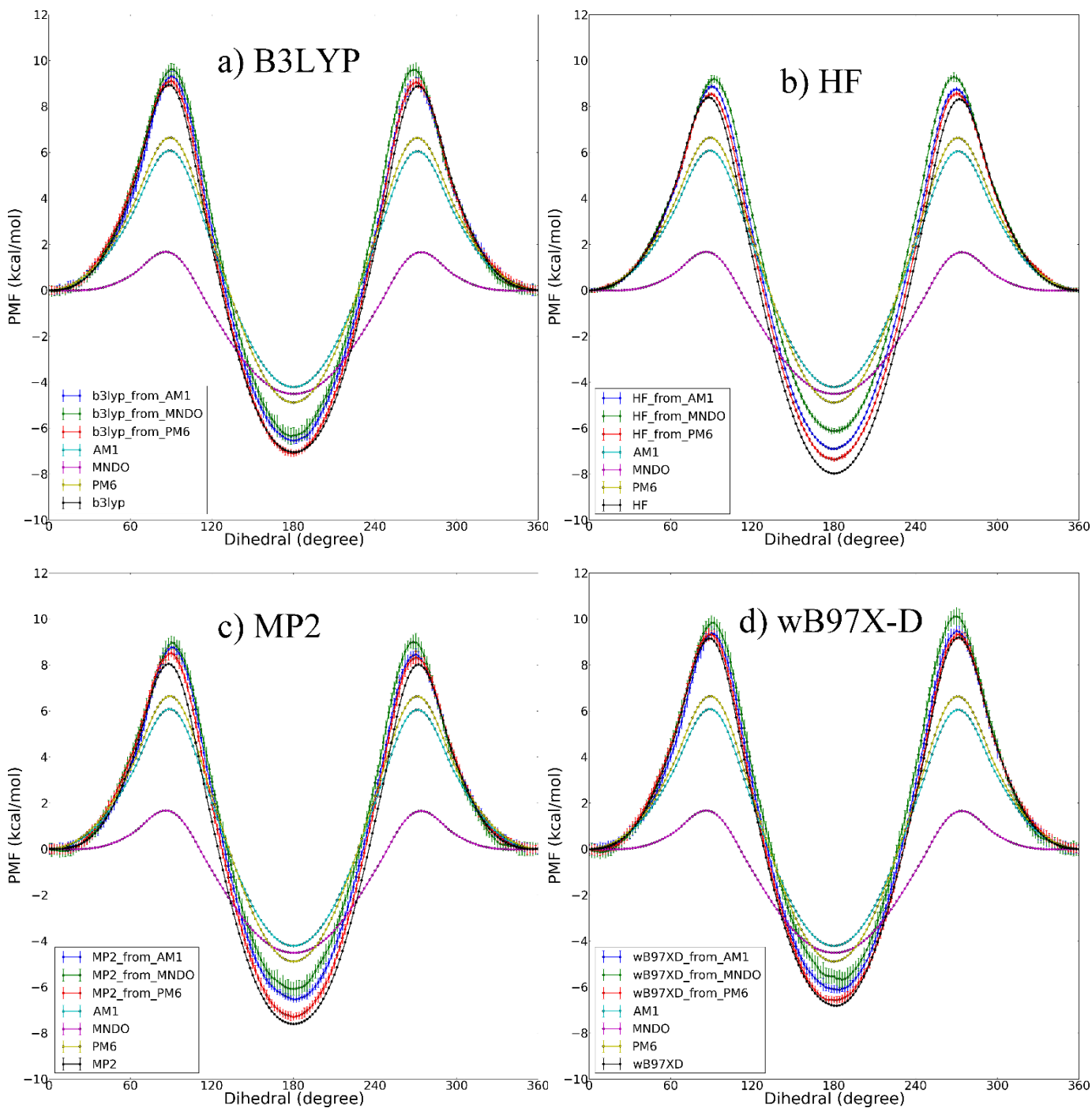


Fig. 5. Comparison between the direct and indirect results at QM levels. a) B3LYP, b) HF, c) MP2, d) ω B97X-D. The error bars are obtained by the normal error propagation procedure.



**Supporting Information: BAR-based Multi-dimensional Nonequilibrium Pulling
for Indirect Construction of QM/MM Free Energy Landscape: From
Semi-empirical to Ab Initio**

Zhaoxi Sun^{1,2*}

*¹State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China
Normal University, Shanghai 200062, China*

²Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany

*To whom correspondence should be addressed: z.sun@fz-juelich.de

Fig. S1. a) σ as a function of force constant in AMBER at 300 K in the dihedral case. b) σ varies in the spaces of force constant in AMBER and simulation temperature in the dihedral case.

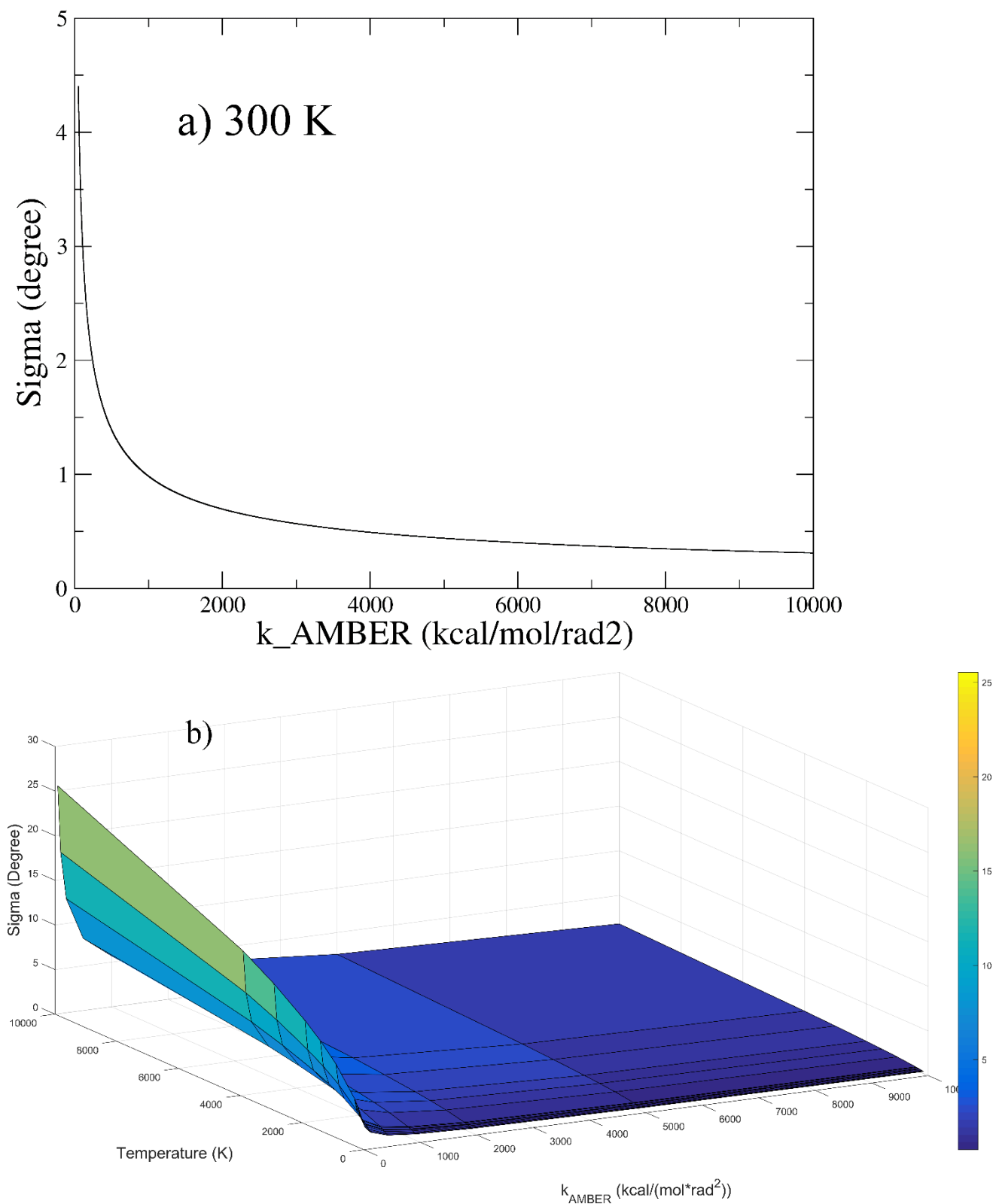


Fig. S2. a) σ as a function of force constant in AMBER at 300 K in the distance case. b) σ varies in the spaces of force constant in AMBER and simulation temperature in the distance case.

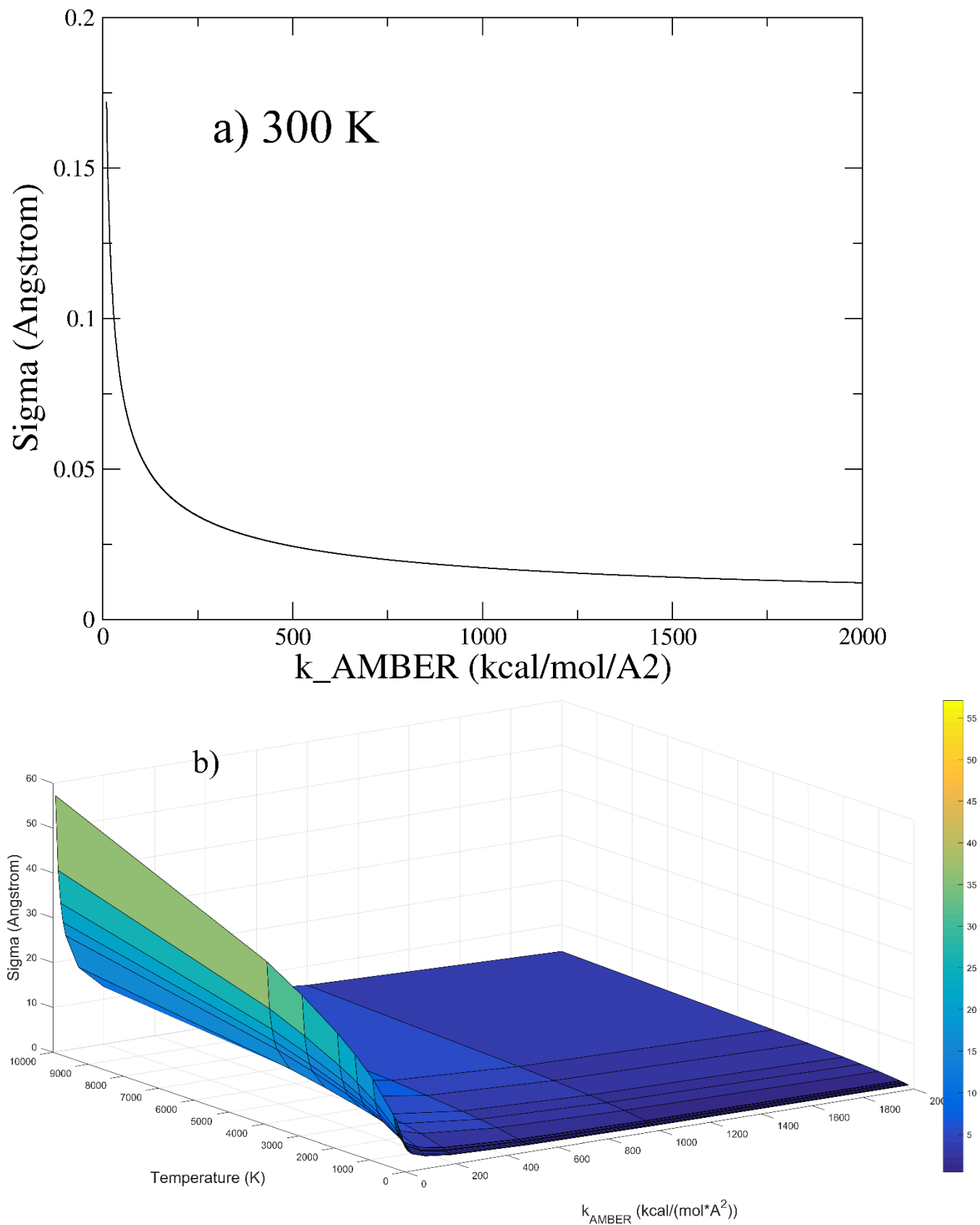


Fig. S3. Under different ab initio QM Hamiltonians, the convergence behavior of free energy profiles the pulling speeds at a) HF, b) MP2, c) ω B97X-D. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. The time in the legend represents the pulling time for each 2° segment. From the pulling-speed dependence, we know that 0.5 ps per segment is slow enough for convergence and thus the statistics under this pulling speed is used for discussion in the following parts of the paper.

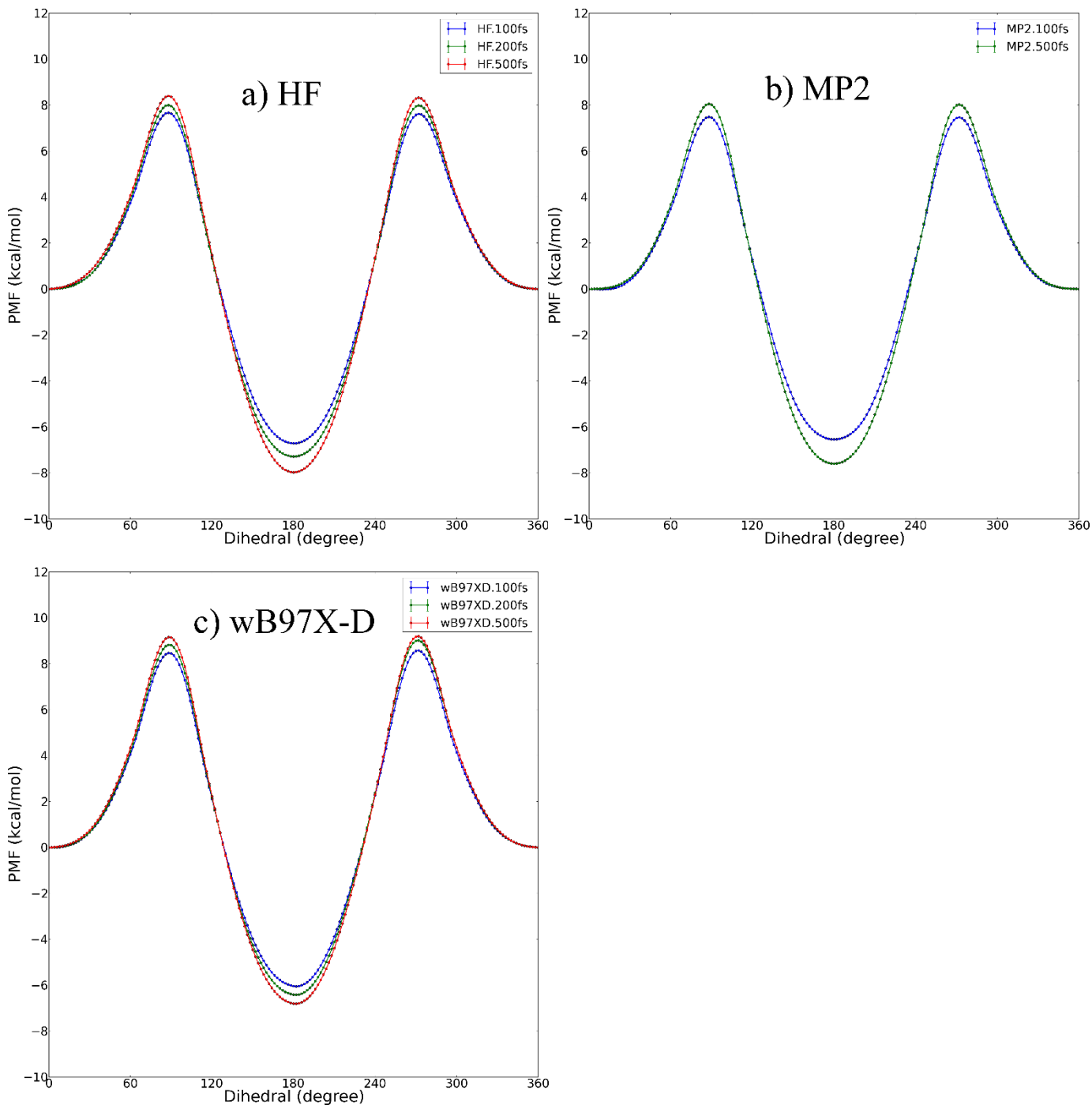
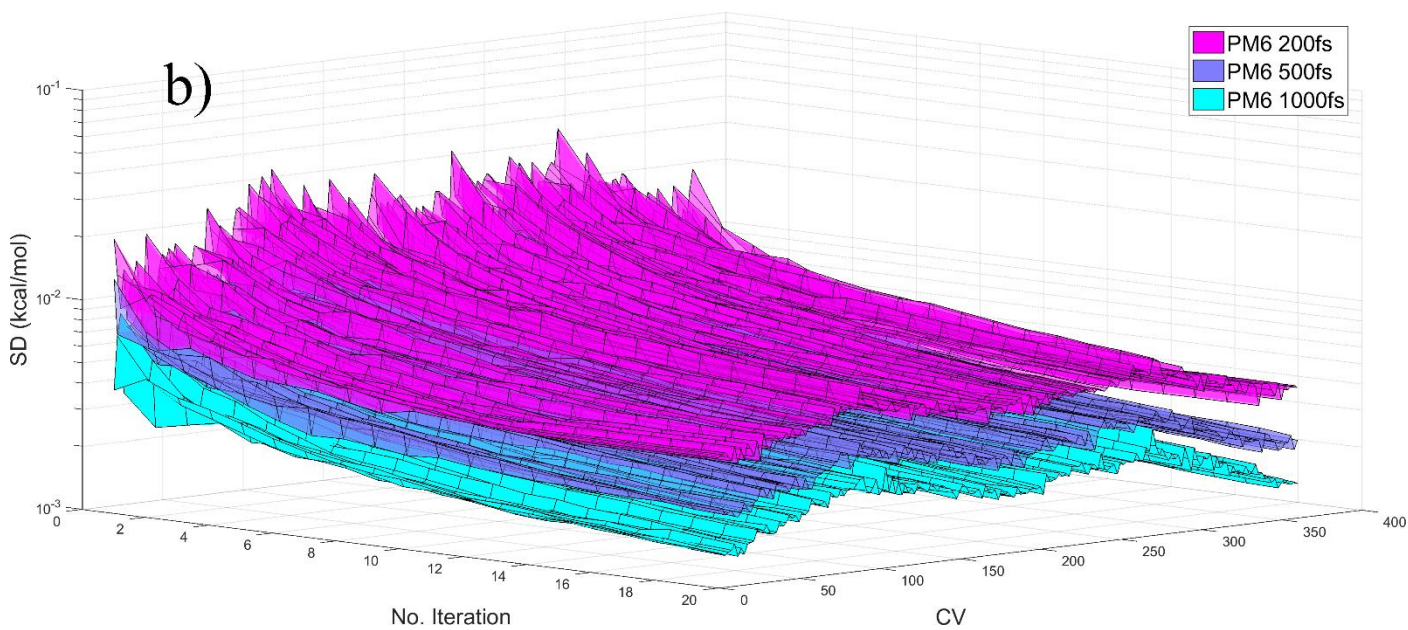
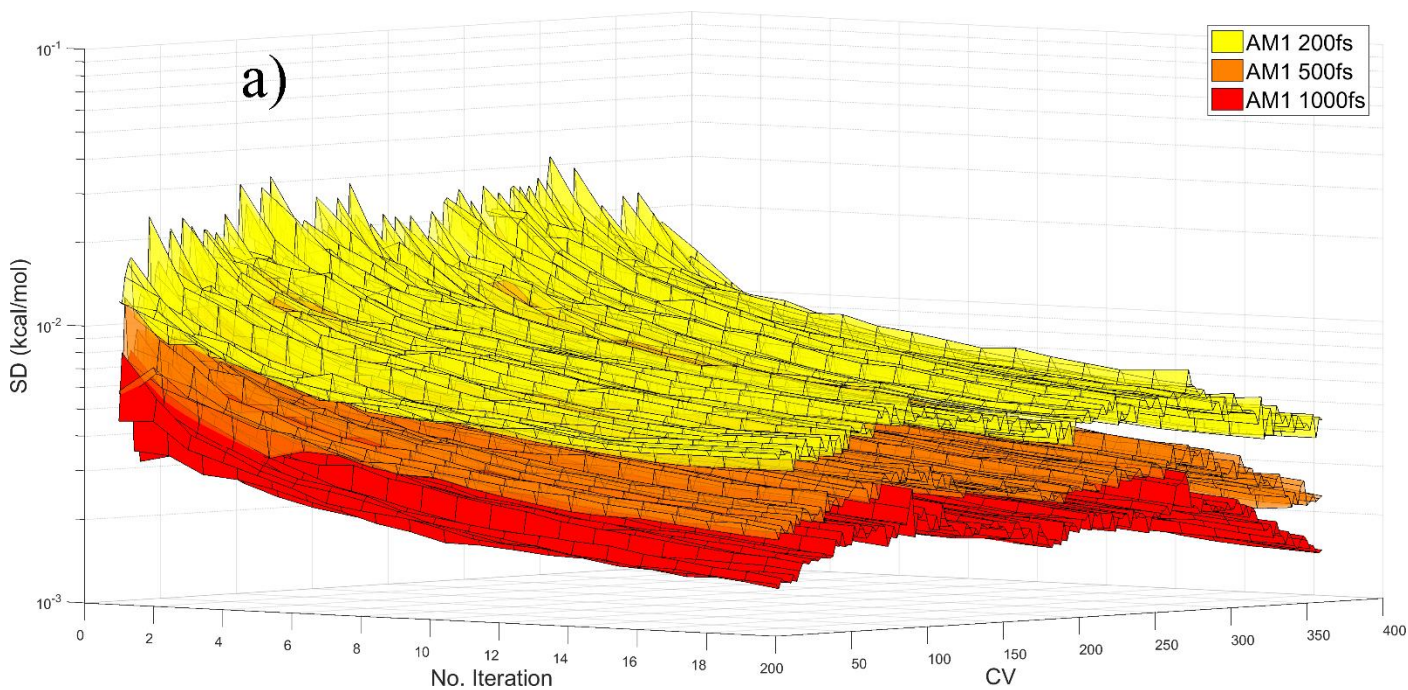


Fig. S4. Sample-size dependence of SD profiles for direct free energy simulation at different levels of theory.

a) AM1, b) PM6, c) MNDO, and d) B3LYP.



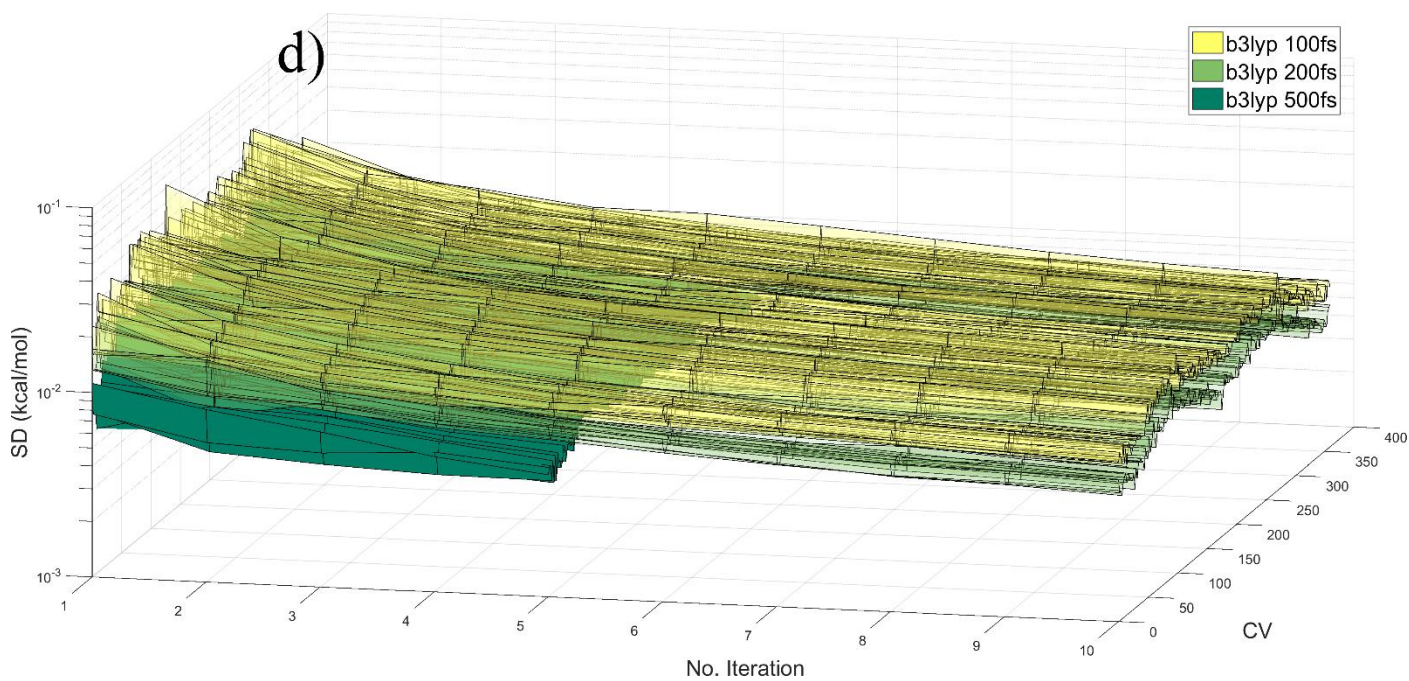
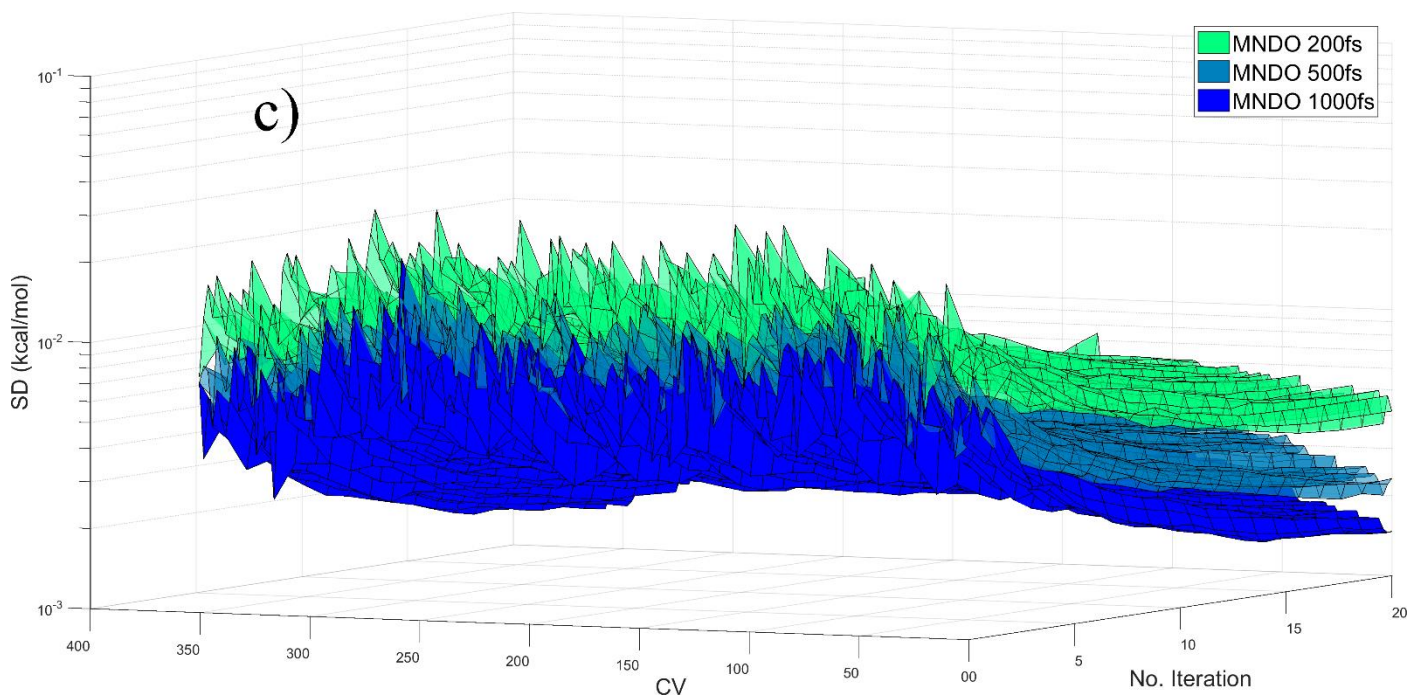


Table S1. Top: the computational times used in the construction of free energy profiles at SQM and ab initio QM levels under the pulling speed of 500 fs per segment. In the direct nonequilibrium free energy simulations under each Hamiltonian, we have 2 directions (forward and backward pulling), 180 segments (from 0°-360° with 2° increments), 50 nonequilibrium realizations per segment. As a result, there are 18000 pulling simulations per system. The Core Time is calculated as 500 fs/segment * 18000 segment / 1,000,000 ns/fs / speed(ns/day) in Table 1.

Bottom: the computational times used in the SQM<->QM corrections under the pulling speed of 2 fs per segment. We have 2 directions (forward SQM-to-QM and backward QM-to-SQM pulling), 180 segments (from 0°-360° with 2° increments), 50 nonequilibrium realizations per segment. As a result, there are 18000 pulling simulations per system. The Core Time is calculated as 1 fs/segment * 18000 segment / 1,000,000 ns/fs / speed(ns/day) in Table 1.

Note that the initial sampling is not included in this Table.

Direct Free Energy Simulations							
Level of Theory Terms	semi-empirical QM				ab initio QM		
	AM1	PM6	MNDO	HF	B3LYP	MP2	wB97XD
Core Time (days)	0.19	0.19	0.19	647.16	3123.92	1390.61	4741.83
Core Time (h)	4.6	4.7	4.6	15531.7	74974.0	33374.5	113804.0
Total (core-h)	237698.2						

SQM<->QM corrections												
SQM to QM Terms	AM1 to				PM6 to				MNDO to			
	HF	B3LYP	MP2	wB97XD	HF	B3LYP	MP2	wB97XD	HF	B3LYP	MP2	wB97XD
Core Time (days)	1.29	6.25	2.78	9.48	1.29	6.25	2.78	9.48367	1.29431	6.247831	2.78121	9.48367
Core Time (h)	31.1	149.9	66.7	227.6	31.1	149.9	66.7	227.6	31.1	149.9	66.7	227.6
Total (core-h)	1426.1											

Table S2. Relative free energy in kcal/mol at important points on the free energy profiles at ab initio QM levels obtained from direct and indirect schemes.

Scheme \ Position	b3lyp from				HF from				MP2 from				wB97X-D from			
	direct	AM1	PM6	MNDO	direct	AM1	PM6	MNDO	direct	AM1	PM6	MNDO	direct	AM1	PM6	MNDO
minimum at 0°	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
peak at 90°	8.9	9.2	9.1	9.5	8.4	8.8	8.5	9.1	8.0	8.7	8.5	8.8	9.2	9.3	9.3	9.7
minimum at 180°	-7.0	-6.5	-7.1	-6.4	-8.0	-6.9	-7.3	-6.1	-7.6	-6.5	-7.3	-6.1	-6.8	-6.1	-6.6	-5.5

Table S3. Efficiency comparison of direct and indirect free energy simulation at QM level. Total simulation time in direct scheme is given by $N_{\text{segments}} * N_{\text{traj}} * (\phi_{\text{NEW}} + \phi_{\text{eq}})$, while the total simulation time in the indirect scheme is the sum of $N_{\text{segments,SQM}} * N_{\text{traj,SQM}} * (\phi_{\text{NEW,SQM}} + \phi_{\text{eq,SQM}})$ at SQM level and $N_{\text{traj,SQM}\rightarrow\text{QM}} * (\phi_{\text{NEW,SQM}\rightarrow\text{QM}} + \phi_{\text{eq,SQM}}) + N_{\text{traj,QM}\rightarrow\text{SQM}} * (\phi_{\text{NEW,QM}\rightarrow\text{SQM}} + \phi_{\text{eq,QM}})$ in SQM \leftrightarrow QM correction. N_{segments} is the number of segments and N_{traj} is the number of realizations per segment. The simulation time at QM level is scaled by the ratio of computational cost under QM Hamiltonian and that under SQM Hamiltonian in Table 1 to be the effective simulation time at SQM level, enabling direct comparison between computational costs. The computational cost of SQM \rightarrow QM differs from QM \rightarrow SQM, as the initial configuration sampling procedures proceed under different Hamiltonians. The computational-cost comparisons for all QM Hamiltonians reported in this work except B3LYP are included.

Terms Simulation	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ps) scaled to SQM Hamiltonian	Relative efficiency
direct SQM	0.05	0.5x2=1	180	20	3780.00	3343.64
SQM \rightarrow HF	0.05	0.001	180	20	12217.10	-
HF \rightarrow SQM	0.05	0.001	180	20	613892.28	-
indirect HF	-	-	-	-	629889.39	20.07
direct HF	0.05	0.5x2=1	180	20	12638958.80	1.00

Terms Simulation	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ps) scaled to SQM Hamiltonian	Relative efficiency
direct SQM	0.05	0.5x2=1	180	20	3780.00	7184.80
SQM \rightarrow MP2	0.05	0.001	180	20	26045.27	-
MP2 \rightarrow SQM	0.05	0.001	180	20	1319128.55	-
indirect MP2	-	-	-	-	1348953.82	20.13
direct MP2	0.05	0.5x2=1	180	20	27158529.05	1.00

Terms Simulation	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ps) scaled to SQM Hamiltonian	Relative efficiency
direct SQM	0.05	0.5x2=1	180	20	3780.00	24499.47
SQM->wB97XD	0.05	0.001	180	20	88378.10	-
wB97XD->SQM	0.05	0.001	180	20	4498103.27	-
indirect wB97XD	-	-	-	-	4590261.37	20.17
direct wB97XD	0.05	0.5x2=1	180	20	92608008.43	1.00