JCTC Journal of Chemical Theory and Computation

Article

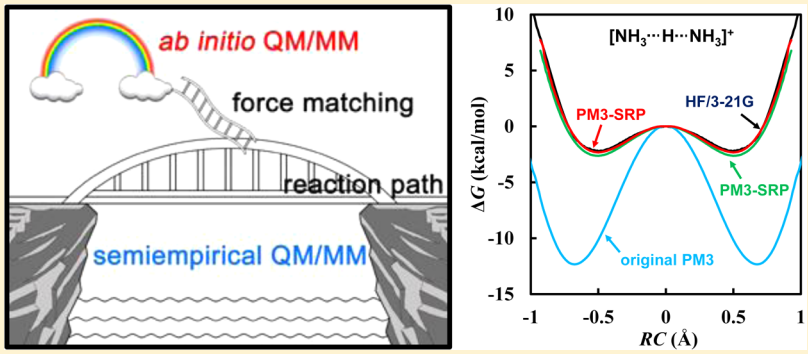
pubs.acs.org/JCTC

Reaction Path Force Matching: A New Strategy of Fitting Specific Reaction Parameters for Semiempirical Methods in Combined QM/MM Simulations

Yan Zhou and Jingzhi Pu*

Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N. Blackford St., LD326, Indianapolis, Indiana 46202, United States

Supporting Information



ABSTRACT: We present a general strategy of reparametrizing semiempirical (SE) methods against *ab initio* (AI) methods for combined quantum mechanical and molecular mechanical (QM/MM) simulations of specific chemical reactions in condensed phases. The resulting approach, designated Reaction Path Force Matching (RP-FM), features cycles of sampling configurations along a reaction path on an efficient SE/MM potential energy surface (PES) and adjusting specific reaction parameters (SRPs) in the SE method such that the atomic forces computed at the target AI/MM level are reproduced. Iterative applications of the RP-FM cycle make possible achieving the accuracy of AI/MM simulations without explicitly sampling the computationally expensive AI/MM PES. The bypassed sampling, nevertheless, is implicitly accomplished through the aid of the efficient SE-SRP/MM PES, on which the target-level reaction path is expected to be obtained upon convergence. We demonstrate the effectiveness of the RP-FM procedure for a symmetric proton transfer reaction in the gas phase and in solution. The remarkable agreements between the RP-FM optimized SE-SRP methods and the target AI method on various properties, including energy profiles, potential of mean force free energy profiles, atomic forces, charge populations, and solvation effects, suggest that RP-FM can be used as an efficient and reliable strategy for simulating condensed-phase chemical reactions.

1. INTRODUCTION

Reliable simulations of free energy profiles for chemical reactions in condensed phases and biological systems require proper sampling of the configuration space based on quantum mechanical potential energy surfaces (PES).[1] Such requirements make combined quantum mechanical and molecular mechanical methods (QM/MM)[2−7] an especially appealing approach. In the QM/MM treatment, a small electronically reactive subsystem (referred to as the QM region) that involves chemical bond breaking and forming is described by quantum mechanics, and the rest of the system (the MM region), often consisting of a large number of nonreactive molecular fragments, is modeled by efficient classical MM force fields. For the choice of the QM level of theory, although *ab initio* molecular orbital (AI-MO)[8] and density functional theory (DFT) methods,[9,10] here collectively referred to as *ab initio* QM (AI) methods, are generally preferred for accuracy and reliability, the computational costs associated with AI calculations preclude AI/MM dynamical simulations to be performed with sufficient statistical sampling, especially for large-sized QM/MM systems. Even aided with extended Lagrangian techniques, such as Car−Parrinello[11,12] and density-matrix propagation[13] molecular dynamics, which eliminate self-consistent field iterations in solving the Schrödinger equation, direct samplings on an AI/MM PES for an extended period of time (e.g., ns and beyond), as conventionally done in pure MM simulations, remain practically infeasible.

On the other hand, semiempirical QM (SE) methods,[14−17] such as MNDO,[18] AM1,[19] PM3,[20] and the related variants with Pairwise Distance Directed Gaussians (PDDG)[21,22] are several orders of magnitude more efficient than the AI methods. The efficiency of the SE methods is rooted in the use of

Received: November 4, 2013
Published: March 26, 2014

   dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

parametrically simplified electronic integrals under approximations such as neglect of diatomic differential overlap (NDDO),[23] together with the use of pseudominimal basis sets.[16] Beyond these so-called MNDO-type methods, versatility and theoretical rigor can be further built into SE methods by treating *d* orbitals[24,25] and including orthogonalization corrections[26,27] explicitly. Due to their cost-effectiveness, SE methods have been widely used in combined QM/MM simulations,[7,16,17] with which a moderate amount of sampling of the reactive systems becomes feasible. For example, one- and two-dimensional umbrella sampling simulations can be routinely carried out at SE/MM levels to yield satisfactory free energy profiles,[28,29] provided that the selected SE method is properly validated against high-level computational benchmarks and/or experiments.

Ideally, however, one would like to perform statistical simulations on a PES that reaches the accuracy of AI/MM but computed as efficiently as SE/MM. In practice, this goal can be achieved, at least in part, by reparametrizing a selected SE method. Such a strategy has proven especially useful when a specific system is concerned. For a specific reaction, if an SE method yields non-negligible errors compared with the benchmark AI results, improvements can be made by obtaining a set of specific reaction parameters (SRPs).[29−33] Optimization of a SE-SRP method typically involves one or more of the following steps: (a) construct a training set of gas-phase molecular models that mimic the actual reactive system; (b) obtain benchmark data, such as reaction energies, barrier heights, vibrational frequencies, atomic charges, dipole moments, etc., from AI calculations (or directly from experiments if even AI methods are inaccurate), for which the geometries of the relevant stationary species (e.g., reactants, products, and transition states) optimized at the corresponding AI levels are often used; (c) adjust the selected parameters (i.e., the SRPs) in the SE method such that the tailored SE-SRP method satisfactorily reproduces the AI benchmark data; and finally, (d) apply the resulting SE-SRP method to gas-phase or condensed-phase dynamical simulations.

The stationary point-based gas-phase SRP fitting strategy, however, is not exempt from drawbacks. Numerically, "overfitting" may occur when only a limited amount of AI data for a few stationary species is used in the parametrization process. From the physical perspective, the rationale of deriving SRPs this way relies on two basic assumptions: (i) The model systems closely resemble the actual chemical system of interest. (ii) The SRPs calibrated in the gas phase are transferrable to the condensed-phase simulations. One should be cautious about possible violation of the first assumption when an attempt is made to apply SRPs obtained for one system to a significantly different chemical situation without careful validation. For example, using monophosphate esters as models to obtain SRPs for reactions of phosphoanhyride, such as adenosine triphosphate (ATP), is under the risk of generating inapplicable parameters. The circumstances under which the "transferability" assumption becomes invalid are discussed next in detail.

The assumption that calibrating SRPs in the gas phase is sufficient for the subsequent condensed-phase simulations can become problematic in a number of situations. First, for certain systems, fitting parameters solely based on gas-phase properties may introduce physical imbalance in the resulting SRP method. We illustrate this issue by using ATP again as an example. For hydrolysis of ATP, the reaction energy in the gas phase is about −168 kcal/mol,[34] in comparison with a free energy of reaction of about −8 kcal/mol in aqueous solution. Therefore, a successful SE-SRP method for studying solution-phase ATP hydrolysis is required to give not only accurate gas-phase reaction energies but also reliable solvation free energy changes. Whether an SE-SRP optimized for gas-phase quantities can be used effectively to study condensed-phase reactions is further complicated by the fact that conventional SE methods are known to underestimate the electronic polarizabilities.[35,36] Consequently, SE-SRPs fit in the gas phase lack the guaranteed transferability when employed in QM/MM simulations in the condensed phases, if the polarization of the QM fragment by the MM electric field is inaccurately described and needs further adjustments. Second, for systems such as highly charged phosphate anions, the related gas-phase species are unstable (readily ionized),[37] making them questionable models for reactions in solution or in enzymes. Third, it is known that certain molecules such as pyrophosphate anions[37] and guanosine triphosphates (GTP)[38] can change their conformation significantly when they are transferred from the gas phase into condensed-phase environments. Such conformational distinctions, if exist, make the use of gas-phase models less relevant for obtaining SRPs applicable to the condensed-phase simulations.

Due to these reasons, adjusting SE methods against benchmark AI data in condensed-phase environments, ideally based on an ensemble of configurations sampled consistently at the AI level, is highly desirable for obtaining reliable SRPs. However, the computational costs associated with AI PES calculations make brute force samplings of the condensed-phase systems on such surfaces a formidable task. To resolve the dilemma between reliability and efficiency, here we introduce a new SRP-fitting strategy that aims at achieving sampling on AI/MM PESs at a computational cost reduced by several orders of magnitude. In this strategy, an SE method is reparametrized against a target AI method, such that when the optimized SE-SRP method is employed in QM/MM simulations in collecting configurations along a reaction path, the atomic forces computed at the target AI/MM level on these configurations are faithfully reproduced at the SE-SRP/MM level. As the reaction path and force matching[39−45] techniques are used in combination, we name this strategy as Reaction Path Force Matching (RP-FM). Because of the "self-consistent" nature of the problem, i.e., the force-matching optimization of an SE-SRP method depends on a set of reaction path configurations determined by using the optimized SE-SRP method itself, RP-FM needs to be conducted in an iterative fashion. Upon convergence, the SE-SRP method optimized by RP-FM reaches the accuracy of the target AI method, and the goal of sampling an AI/MM-quality PES at the cost of SE/MM is achieved.

Condensed-phase-based force matching has been proposed to be advantageous in deriving pure MM force field parameters[40,41,43,45,46] and defining reactive force fields.[47] For example, under condensed-phase environments, steric and polarization as well as pressure/temperature effects are implicitly incorporated in the MM parameter fitting process.[43] In addition, the solute conformations accessed in condensed phases are usually more "open" compared with the gas-phase ones, in which formation of intramolecular hydrogen bonds tends to make the solute "closed up".[43] Thus, when condensed-phase configurations are used, information on explicit hydrogen bonding with the solvent molecules are better incorporated in

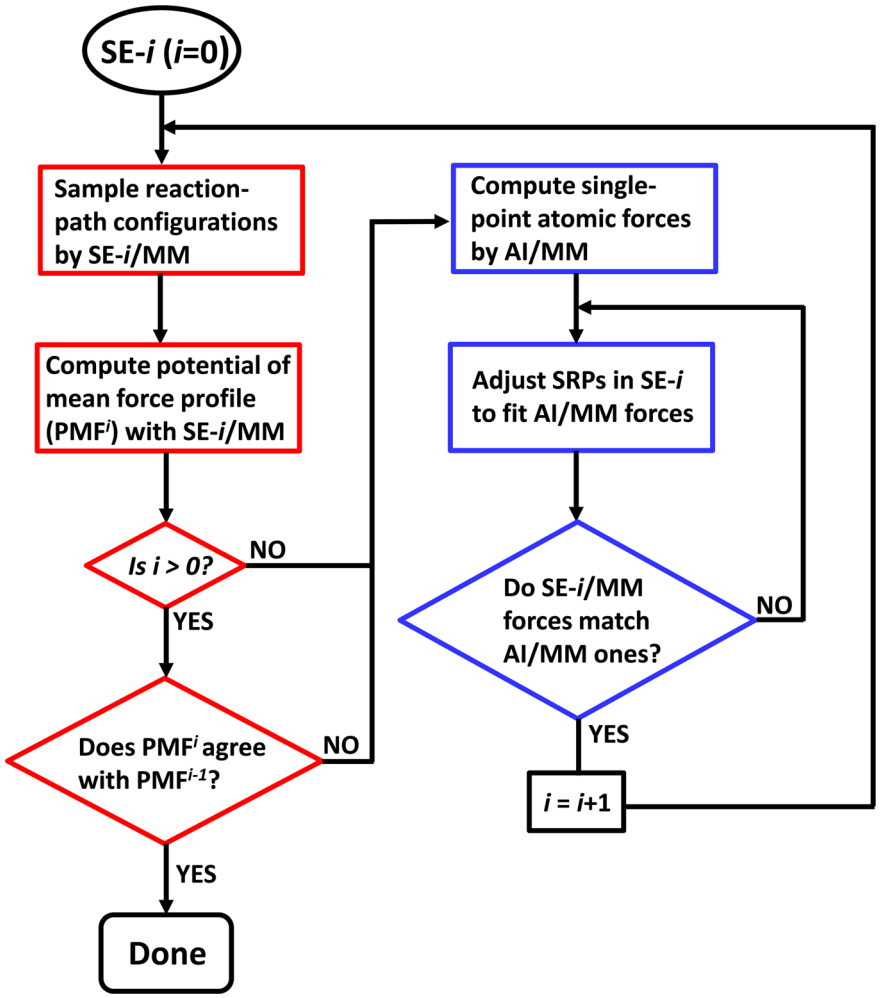


Figure 1. A flowchart representation of the generic procedure for conducting Reaction Path Force Matching (RP-FM) in condensed phases. The RP-FM procedure iterates over cycles, each consisting of two stages: (i) a reaction path sampling (RP) stage, with its constituent steps shown in the red boxes, followed by (ii) a force-matching (FM) stage, with the associated steps enclosed in the blue boxes. The RP stage obtains necessary configurations along a reaction path by using efficient SE/MM simulations. On the basis of the RP configurations, the SE-SRP method is tuned against the AI method in the FM stage to reproduce the target AI/MM atomic forces. In the flowchart, the original SE method is denoted as SE-0, and the SE-SRP method optimized in the *i*th RP-FM iteration as SE-*i*. For condensed-phase RP-FM, the convergence of the SRP fitting is monitored by examining the PMF profiles obtained from SE-*i*/MM simulations over the RP-FM iterations. When applied to gas-phase systems, RP-FM can be carried out following a similar procedure, except that the potential energy surfaces are computed from pure QM calculations and that convergence is monitored by potential energy profiles.

the parametrization.[43] With force matching, many-body effects are implicitly modeled by utilizing a greater amount of information (3*N* forces vs 1 total energy, with *N* being the number of atoms).[40,45] Finally, condensed-phase force matching offers a more direct connection to the potential of mean force or free energy information[40,41,47] compared with the gas-phase fitting strategy, in which calibrations of the MM parameters are primarily focused on potential energies. Despite these advantages, known limitations of the force-matching strategy include that it likely leads to system-specific parameters[43] and that it may suffer from numerical difficulties associated with nonlinear fitting.[40] Because the focus of the present development is to obtain SRPs for QM/MM simulations of a specific chemical reaction, whether the resulting SE-SRP method can also serve as a general purpose method is much less of a concern. As we will discuss, the nonlinear optimizations of the SRPs are facilitated by using a genetic algorithm.

The rest of the article is organized as follows. In Section 2, we first discuss the general aspects of the RP-FM strategy and offer a generic procedure for implementing the strategy; the related notations are also introduced. In Section 3, we describe the tests we performed to demonstrate the effectiveness of the strategy. We first tested it for a gas-phase proton transfer reaction, and then we applied RP-FM to optimize SE-SRPs for the same reaction in solution. Section 4 contains computational details. Results and discussion are presented in Section 5. In Section 6, we discuss the relation of the RP-MF approach to other work. Concluding remarks are given in Section 6, where an outlook for future development is also provided.

2. GENERAL DESCRIPTION OF RP-FM

2.1. Overall Strategy. The RP-FM strategy begins with a basic assumption that there is a reasonable SE level, either an SE method parametrized for a general purpose or a special-purpose SE method that has been developed in a specific chemical context based on which reparametrization can be

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

made. The efficiency of the SE method makes it affordable to sample condensed-phase reactive systems by QM/MM simulations. Because free energy or potential of mean force (PMF) profiles along selected reaction coordinates are of central importance in QM/MM simulations of chemical reaction mechanisms, the RP-FM strategy is first designed to collect an ensemble of configurations along a reaction path (RP) that connects the reactant and product regions; examples of such reaction paths, in a general sense, include minimum energy paths (MEP) (or paths along the intrinsic reaction coordinate IRC),[48,49] paths defined by a few distinguished reaction coordinates (DRC),[50] minimum free energy paths (MFEP) obtained by the finite temperature string[51,52] or other related methods,[53,54] and reaction paths defined by the energy gap between valence bond[55] states or diabatic states in electron transfer.[56] For convenience, we collectively define the computations to determine or sample configurations along the reaction path as the RP stage. On the basis of the ensemble of configurations sampled on the efficient SE/MM PES, benchmark AI/MM single-point calculations are then performed, taking the process to the next stage, in which force matching is carried out.

The second designed element in the RP-FM strategy is to calibrate the SE method against AI results in terms of atomic forces instead of energies by employing the force-matching technique. We refer to this calibration process as the force-matching (FM) stage. Specifically, in the FM stage, a set of SE-SRPs are adjusted to reproduce the AI/MM forces on selected atoms or groups. Force matching can be conducted conveniently in Cartesian coordinates or alternatively in other coordinate systems such as curvilinear internal coordinates.[57,58] If atomic forces are used, the atoms that are subject to force matching may include but are not limited to the QM atoms, which typically contain the solute molecules for solution-phase reactions or the substrate/cofactor molecules, plus a few important proteins residues when an enzyme active site is studied.

Once the tailored SRPs are obtained to satisfactorily reproduce the target AI/MM forces (along a specific reaction path), the resulting SE-SRP method is used to resample the QM/MM system in the condensed phase to collect an updated pool of reaction path configurations for future force matching. The "prediction/correction" cycle consisting of the RP stage (predictor) and the FM stage (corrector) is then repeated iteratively until convergence is established. Such convergence can be monitored, for example, from the numerical values of SRPs or other reaction path-related properties. As free energy information along the reaction path are readily available, whether the SE-SRP methods converge to give a stable performance can be monitored by examining the potential of mean force (PMF) profiles over RP-FM iterations. Finally, upon completion of RP-FM optimizations, the resulting SE-SRP method can be employed in QM/MM simulations for an extended period of simulation time to examine reactive properties of the specific system or to study other similar systems (for the latter case, transferability needs to be verified). As a result, the QM/MM simulations based on the optimized SE-SRP method are expected to (i) generate an ensemble of configurations as if one directly samples along the reaction path on the target AI/MM PES (a detailed analysis of RP-FM's capability in reproducing the target AI-level reaction path is provided in Section 5.1.1, when the related results are discussed) and (ii) accurately reproduce the AI/MM forces for this reaction-path ensemble such that the free energy profile at the target AI/MM level is faithfully restored.

2.2. Generic Procedure. A generic procedure for implementing the RP-FM strategy in the context of QM/MM simulations is illustrated in Figure 1. In the flowchart, the steps associated with the RP and FM stages are shown in the red and blue boxes, respectively. For convenience, we define the union of a RP stage and the FM stage that immediately follows as a single RP-FM cycle, the repetitions of which form RP-FM iterations. The initial SE method, usually the original SE method, is referred to as SE-0, for which a token iteration, i.e., "iteration 0", is assigned to remind us that no reparametrization has been done yet. Each time when a single RP-FM cycle is completed, the iteration number is increased by 1. The SE-SRP method obtained at the end of the FM stage in the ith RP-FM iteration is labeled as SE-i, which is in turn used to determine the updated reaction path configurations in the RP stage of the $(i+1)$th RP-FM iteration. We also refer to the reaction path sampled consistently on the SE-i/MM PES as the SE-i path. Note that due to the sequential nature of the RP and FM stages in each cycle, the acquisition of the reaction path configurations at a certain SE-SRP/MM level always lags behind the determination of that SE-SRP method by 1 iteration, i.e., the reaction path configurations used throughout the ith iteration are based on the SE-$(i-1)$ parameter set, which is obtained from the $(i-1)$th iteration. Following the same convention, the PMF profile determined from the SE-i/MM simulations is labeled as PMF^i, which is used to monitor convergence of the RP-FM iterations.

Note that although the above procedure is presented for condensed-phase simulations in the combined QM/MM framework, the same procedure can be adopted for gas-phase systems, except that potential energies will be computed by pure QM calculations and convergence will be monitored by using potential energy profiles.

3. TEST CASES

We tested the RP-FM approach systematically at two levels. At the first level, we examined the feasibility of the method by carrying out proof-of-concept RP-FM calculations based on a gas-phase reaction path. At the second level, we applied the RP-FM procedure in solution, where the configurations for force matching were directly sampled from SE/MM simulations along a free energy pathway.

In particular, we performed tests at both levels for a symmetric proton transfer reaction between ammonium and ammonia (reaction 1)

$$NH_4^+ + NH_3 \rightarrow NH_3 + NH_4^+ \quad (1)$$

The PM3[20] and Hartree−Fock (HF)[59] methods were chosen as the SE and target AI level, respectively. We used this combination because the PM3 Hamiltonian has been shown to overestimate the gas-phase barrier height of reaction 1 by about 9 kcal/mol compared with the HF/3-21G[60] results.[61,62] Therefore, directly employing the PM3 method in solution-phase QM/MM simulations of reaction 1 would lead to free energy profiles that significantly differ from the HF/3-21G/MM ones. We will demonstrate how these discrepancies can be systematically removed by using the RP-FM approach.

4. COMPUTATIONAL DETAILS

This section contains the detailed descriptions of the simulations outlined previously for testing the RP-FM procedure in the gas phase and in solution. Aspects common to these simulations, including the solute/solvent models, computation of the potential energies, definition of the reaction coordinate, and the force-matching engine, are described in Sections 4.1−4.4. Specific details associated with the gas-phase RP-FM calculations are given in Section 4.5. Additional details for performing the RP-FM simulations in solution are provided in Sections 4.6−4.7. Simulation protocols relevant to analyzing the solution-phase RP-FM results are presented in Sections 4.8 and 4.9.

4.1. Description of Solute Model. The topologies for the solute molecules NH_4^+ and NH_3 were built based on similar protein residues available in the standard CHARMM topology files. The atom types NH3 and HC were used for nitrogen and hydrogen in NH_4^+, similar to the ε-amino group in lysine, whereas the corresponding atom types were set to NH2 and H for NH_3, as in the amine group in glutamine. With the atom types specified, van der Waals (vdW) parameters were assigned based on the standard CHARMM22 force field;[63] these parameters are required for computing the nonbonded vdW interactions between the QM and MM atoms in a combined QM/MM treatment of reaction 1 in solution (see below).

4.2. Potential Energy Calculations. For the gas-phase calculations of reaction 1, the whole system is treated quantum mechanically by the PM3, PM3-SRP, or HF/3-21G methods. For a QM/MM treatment of reaction 1 in aqueous solution, the solute molecules consisting of NH_4^+ and NH_3 are treated by QM at the PM3, PM3-SRP, or HF/3-21G levels, whereas the solvent molecules represented by the modified TIP3P model[64] are treated by MM. The MNDO97[65] package incorporated into the CHARMM[64,66] program (version c35a2) was used for the PM3(-SRP) and PM3(-SRP)/MM calculations. The Q-Chem[67] package (version 4.0.1) interfaced with CHARMM was used for the HF/3-21G and HF/3-21G/MM calculations.

4.3. Definition of Reaction Coordinate. Let N_+ and N represent the nitrogen atoms in the NH_4^+ and NH_3 groups, respectively, and H the transferred proton. To describe the progress of proton transfer for reaction 1, we adopted a distinguished reaction coordinate (RC), which is defined as the antisymmetric combination of the breaking (N_+−H) and forming (N−H) bond distances

$$\mathrm{RC} = r(\mathrm{N_+}\text{−}\mathrm{H}) - r(\mathrm{N}\text{−}\mathrm{H}) \tag{2}$$

This reaction coordinate was used in obtaining both the gas-phase and solution-phase reaction paths.

4.4. Force-Matching Optimization of SRPs Using a Genetic Algorithm. For the force-matching stage in each RP-FM cycle, a micro-genetic algorithm (μGA)[68] was employed to optimize the selected SE parameters. In particular, for reaction 1, a set of 21 parameters in PM3, including 12 parameters for nitrogen and 9 parameters for hydrogen, was chosen to be optimized (see Table S1 of the Supporting Information for a list of the PM3 parameters that were subject to change). The use of μGA in combination with an elitism algorithm made possible an efficient exploration of the relatively large parameter space with a small number of populations[68] (eight populations were used in the present study). To achieve stable optimizations, the PM3 parameters under adjustment were only allowed to change within ±5% of their original values in the standard PM3 method; other parameters excluded from the SRP optimizations were fixed at their original values.

To reproduce the target atomic forces obtained at the HF/3-21G(/MM) level, the μGA driver optimizes the PM3-SRP methods to maximize a fitness function χ, which is defined as the negative magnitude of the root-mean-square-deviation (RMSD) between the forces computed at the two levels

$$\chi = -\sqrt{\frac{1}{3 \times L \times N}\sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{\alpha=x,y,z}\left[f_{ij\alpha}^{\text{PM3-SRP(/MM)}} - f_{ij\alpha}^{\text{HF/3-21G(/MM)}}\right]^2} \tag{3}$$

where L represents the number of configurations included in the force-matching stage, N is the number of atoms on which the force deviations are computed, and $f_{ij\alpha}$ represents the α component of the Cartesian forces on the jth atom in the ith configuration.

4.5. RP-FM along Gas-Phase Reaction Paths. A gas-phase reaction path of reaction 1 was first determined on the original PM3 (PM3-0) surface, along the reaction coordinate RC defined in eq 2. In practice, we generated this initial reaction path (and the updated ones over RP-FM iterations) by using a coordinate driven procedure, in which restrained geometry optimizations were carried out to scan the PM3(-SRP) PESs along RC in the range of 0.0−0.8 Å with a stepsize of 0.1 Å. We note that although one can determine an minimum energy path (MEP) more rigorously by following the intrinsic reaction coordinate (IRC), the distinguished reaction path obtained here can be viewed as an approximated MEP. For convenience of discussion, we simply refer to the resulting gas-phase reaction path as the MEP. Because the proton transfer we consider here is a symmetric reaction, only the reactant branch of the path that connects the reactant state (R) and the transition state (TS) regions was explicitly constructed.

The PES scans described above resulted in MEPs each consisting of nine configurations. Initially, based on the reaction path configurations determined at the PM3-0 level, single-point energies (SPE) and the associated atomic forces (thus we also refer to these single-point calculations as SPEF calculations) were computed at the target AI level, i.e., HF/3-21G. For the gas-phase system of reaction 1 that contains nine atoms, each set of the SPEF calculations generated a total of 243 Cartesian force components at the HF/3-21G level, which were used as the target forces to fit in the FM stages.

The goal of the FM stage in each RP-FM cycle is to adjust SRPs in the SE method such that the forces produced from the adjusted SE-SRP method closely match their target values obtained at the AI level. For the gas-phase study, 21 PM3 parameters (Tables S1 and S2, Supporting Information) were to be optimized to reproduce 243 Cartesian force components at the target HF/3-21G level. These nonlinear optimization processes were managed by the μGA[68] driver described in Section 4.4. In each RP-FM cycle, the FM stage was paired with a fixed reaction path; only after force matching is satisfactorily achieved can one proceed to the RP stage in the next iteration to update the reaction path. Cycles of RP and FM were then repeated iteratively by following the generic procedure presented in Section 2.2. In the present study, we carried out five gas-phase RP-FM iterations, which seem sufficient in obtaining converged energy profiles.

Regarding computational loads, in the FM stage of each cycle, the μGA driver repeatedly invoked MNDO97 calcu-

lations (through CHARMM) to compute the atomic forces with the attempted PM3-SRPs on the same nine reaction-path configurations until the fitness function χ (eq 3, with $L = N = 9$) was maximized, whereas the demanding force calculations at the HF/3-21G level were only conducted once for these configurations; the target force values, once obtained, remained fixed within each FM stage.

4.6. PMF Simulations by Umbrella Sampling. To represent reaction 1 in aqueous solution, the solute molecules NH_4^+ and NH_3 were solvated by a cubic water box of 40 Å × 40 Å × 40 Å. A chloride anion was added to the system to maintain the overall charge neutrality. The solvated system was sampled by molecular dynamics (MD) simulations based on combined QM/MM potentials. The solutes were treated by QM, whereas the water molecules and the counterion were treated by MM. Periodic boundary conditions were employed by using the CRYSTAL facility in CHARMM.[64,66] To smoothly attenuate the nonbonded interactions at a cutoff distance of 13 Å, a force switching function was applied on both electrostatic and van der Waals interactions between 12 and 13 Å. The parameters CUTNB and CUTIM, used for generating the nonbonded atom lists involving the primary system and periodic lattice images, were set to 14 and 15 Å, respectively. The nonbonded atom lists were constructed using a group-based scheme and updated heuristically during the MD simulations. Constant pressure and temperature (NPT) MD simulations were performed at 1 atm and 300 K by using a Nosé−Hoover thermostat.[69,70] The leapfrog Verlet algorithm[71] was used to integrate the equations of motion with a time step of 1 fs.

To facilitate convergence of sampling, we harmonically anchored the N_+−N bond distance at 2.7 Å. The N_+−H−N angle was restrained similarly at the linear configuration. The spring force constants for the distance and angle restraints were set to 300 kcal/mol/Å^2 and 300 kcal/mol/rad^2, respectively. In addition, the center of mass of the two nitrogen atoms was tethered to the center of the water box with a harmonic force constant of 200 kcal/mol/Å^2. During the MD simulations, the internal geometries of TIP3P water molecules were constrained by using the SHAKE algorithm.[72]

To apply the RP-FM procedure on reaction 1 in solution, umbrella sampling[73] simulations were carried out for collecting statistical samples along a free energy path represented in the reaction coordinate RC (see eq 2 for definition). Specifically, we distributed seven windows along RC in the range of −1.0 to 0.2 Å, with their centers placed at an interval of 0.2 Å. For each umbrella window sampled by PM3(-SRP)/MM, we performed 100 ps MD simulations for equilibration, followed by additional simulations of 200 ps for data collection. To establish the benchmark PMF profiles, separate umbrella sampling simulations were conducted at the target HF/3-21G/MM level. Given the computational demands associated with the HF/3-21G/MM calculations, shorter samplings consisting of 15 ps equilibration and 30 ps data collection were used for each umbrella window in these benchmark AI/MM MD simulations. Finally, all PMF profiles were determined by the weighted histogram analysis method (WHAM).[74]

4.7. Force Matching along Free Energy Paths in Solution. In the initial RP stage, umbrella sampling simulations for reaction 1 using the original PM3 method (PM3-0) was first carried out following the simulation protocols described in Section 4.6. Instantaneous configurations sampled in each umbrella window during the 200 ps MD trajectories in the data-production runs were saved every 10 ps, resulting in a collection of 20 frames, among which 10 frames (every other frame starting from the first frame) were selected for HF/3-21G/MM SPEF calculations. Hence, for the seven umbrella windows that cover the reactant branch of the reaction path, 70 configurations were selected for force matching. In the present work, only the forces on the QM atoms, i.e., the solute molecules NH_4^+ and NH_3, were included in force matching. Therefore, for the nine-atom solute system of reaction 1, each of the 70 sets of Cartesian forces is composed of 27 components; as a result, a total of 1,890 Cartesian force components were explicitly considered in the solution-phase RP-FM, for which the fitness function to optimize in the FM stages adopts the functional form of eq 3, with $L = 70$ and $N = 9$. We note that although the forces on the MM atoms do not explicitly appear in the fitness function, they do contribute to determination of SRPs in the solution-phase RP-FM, given the fact that the forces on the QM atoms are computed in the presence of MM solvent molecules.

To maximize the fitness function (or equivalently to minimize the deviations between the PM3-SRP/MM and HF/3-21G/MM forces), we employed the μGA driver described in Section 4.4 to adjust the 21 selected parameters in PM3 (Table S1, Supporting Information) by invoking necessary PM3-SRP/MM SPEF calculations in each FM stage. Once the force deviation is satisfactorily minimized, one completes the FM stage and moves to the RP sampling stage for another iteration. By using the newly determined PM3-SRP parameters, another set of QM/MM umbrella sampling simulations are initiated to obtain an updated ensemble of configurations along the reaction path. The RP-FM cycle consisting of umbrella sampling along a reaction path, and force matching is then repeated following the procedure described in Section 2.2 and Figure 1, until the PMF free energy profiles determined from the PM3-SRP/MM methods display stable converging behavior. In the present work, five RP-FM iterations were conducted in solution to converge the PMF profiles for reaction 1.

4.8. Mulliken Charge Calculations. We performed charge analyses for the solution-phase QM/MM simulations. Mulliken charges[75] on the QM atoms were computed based on QM/MM MD sampled individual configurations. We computed the Mulliken charges on the solute fragments, i.e., NH_4^+ and NH_3, for a total of 140 configurations (20 frames in each of the seven umbrella windows) sampled along the solution-phase free energy path. The average Mulliken charges at a given reaction coordinate position were collected over configurations binned using their instantaneous RC values with a bin width of 0.1 Å.

4.9. Solute−Solvent Interaction Energy Calculations. To monitor whether the present force-matching scheme, i.e., fitting the total force on each atom, can strike a balance to improve both the gas-phase and the solvation contributions to the free energy barriers, we also characterized the solvation effects based on the MD trajectories sampled from the SE(-SRP)/MM simulations. In particular, we estimated the contribution of solvation by using the total solute (S)−solvent (X) interaction energy (E_{SX}), which is defined as the difference in the electronic energy of the solute molecules in the presence and absence of solvent polarization,[76] computed at the same solute configuration sampled in the solution phase

$$E_{SX} = E_{tot} - E_{gas} \quad (4)$$

with E_{tot} and E_{gas} defined as

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

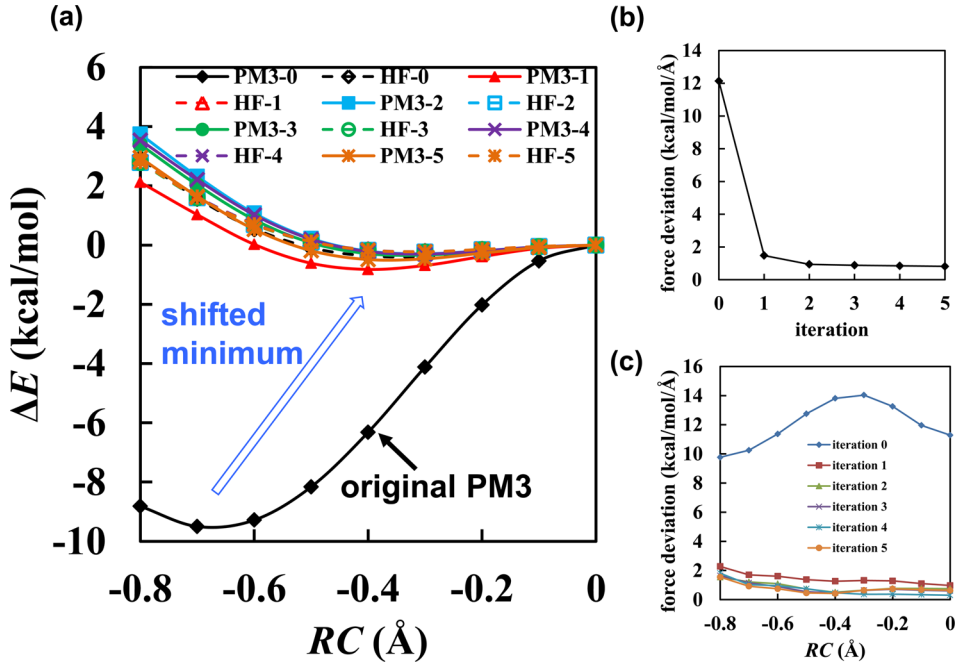


Figure 2. (a) Potential energy profiles for reaction 1 along the reaction coordinate (RC), obtained from the gas-phase RP-FM iterations. PM3-0 denotes the original PM3 method. PM3-*i* (*i* = 1−5) denotes the PM3-SRP method obtained at the end of the FM stage in the *i*th RP-FM iteration. HF-*i* (*i* = 0−5) denotes the HF/3-21G single-point energy profile based on the reaction path determined on the PM3-*i* PES (i.e., HF/3-21G // PM3-*i*). RP-FM successfully shifts the position of the reactant minimum from RC = −0.7 Å (PM3-0) to RC = −0.4 Å (PM3-5), in agreement with the reactant position of RC = −0.3 Å determined at the target HF/3-21G level. (b) Initial force deviations (averaged over 243 force components on nine configurations along each reaction path) between the PM3-SRP and HF/3-21G methods as a function of iteration number. (c) Decompositions of the initial force deviations to individual reaction-path configurations over five RP-FM iterations.

$$E_{tot} = \langle \psi | \hat{\mathbf{H}}_{QM} + \hat{\mathbf{H}}_{QM/MM} | \psi \rangle \quad (5)$$

$$E_{gas} = \langle \psi_0 | \hat{\mathbf{H}}_{QM} | \psi_0 \rangle \quad (6)$$

where $\hat{\mathbf{H}}_{QM}$ is the gas-phase Hamiltonian for the solute molecules, $\hat{\mathbf{H}}_{QM/MM}$ is the hybrid Hamiltonian for QM/MM interactions that captures the electronic polarization of the solute molecules by the solvent, ψ is the solute wave function that is self-consistently determined in the solution phase, and ψ_0 is the gas-phase solute wave function obtained for $\hat{\mathbf{H}}_{QM}$ alone when solvent molecules are removed.

On the basis of the averaged quantities of E_{gas} and E_{SX}, we estimated the gas-phase solute contribution ($\Delta g^{\ddagger}_{gas}$) and the solvation contribution ($\Delta g^{\ddagger}_{SX}$) to the free energy barriers by

$$\Delta g^{\ddagger}_{gas} = \langle E_{gas} \rangle_{TS} - \langle E_{gas} \rangle_{R} \quad (7)$$

$$\Delta g^{\ddagger}_{SX} = \langle E_{SX} \rangle_{TS} - \langle E_{SX} \rangle_{R} \quad (8)$$

where $\langle\ \rangle_R$ and $\langle\ \rangle_{TS}$ denote the ensemble averages of the corresponding energy terms for the reactant state (R) and transition state (TS), respectively. Note that in eqs 7 and 8, the lowercase *g*'s are used to remind us that these quantities are estimated free energy contributions; these estimates developed for the analysis purpose should not be confused with the free energies (*G*) obtained from potential of mean force (PMF) simulations. Finally, we also computed a similar estimate based on E_{tot}

$$\Delta g^{\ddagger}_{tot} = \langle E_{tot} \rangle_{TS} - \langle E_{tot} \rangle_{R} \quad (9)$$

$\Delta g^{\ddagger}_{tot}$ is used to verify that the free energy barriers estimated this way are comparable to those determined directly from the PMF profiles.

In practice, the ensemble-averaged gas-phase solute energies (E_{gas}), solvation contributions (E_{SX}), and total electronic energies (E_{tot}) from PM3(-SRP)/MM simulations were determined "on the fly" by using the QM/MM energy decomposition tool in the MOPAC[15]-based QUANTUM module in CHARMM. On the basis of 100 ps MD simulations, these averaged energy contributions were obtained over 100,000 configurations sampled for both the reactant and transition-state ensembles. Benchmark results of these contributions were evaluated at the HF/3-21G/MM level by SPE calculations based on the configurations sampled from the PM3(-SRP)/MM simulations. In these benchmark calculations, the solution-phase solute energies $\langle \psi | \hat{\mathbf{H}}_{QM} + \hat{\mathbf{H}}_{QM/MM} | \psi \rangle$ (eq 5) were taken from the electronic part (thus the QM/MM vdW interaction energies were excluded) of the total energies computed at the HF/3-21G/MM level for 20 configurations stored along 200 ps MD trajectories, each sampled for the reactant and transition state, and the gas-phase solute energies $\langle \psi_0 | \hat{\mathbf{H}}_{QM} | \psi_0 \rangle$ (eq 6) were computed by performing pure HF/3-21G calculations with all MM solvent molecules and the counterion removed. Note that although only 20 configurations were used in computing the averaged energy components at the HF/3-21G/MM level, whereas such averages were made on 100,000 configurations for PM3(-SRP)/MM, similar standard deviations were obtained in the two treatments (Section 5.2.3 and Table 2), suggesting that the comparison between them is statistically meaningful.

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

5. RESULTS AND DISCUSSION

5.1. Gas-Phase RP-FM. We first demonstrate that RP-FM can be used to upgrade the performance of the selected SE method (PM3) to an AI quality (HF/3-21G) along reaction paths determined in the gas phase. In this case, the ensemble on which we fit the SE forces against their target AI values consists of configurations along an approximate MEP (see Section 4.5 for details), thus can be referred to as an MEP ensemble. As no thermal contributions are taken into account in the MEP ensemble, RP-FM along the gas-phase MEP can be considered as the "zero-temperature" version of the procedure.

5.1.1. Potential Energy Profiles. The potential energy profiles along the MEPs obtained over five RP-FM iterations are shown in Figure 2a. To compare these potential energy profiles, we chose the transition state (TS), which is always located at RC = 0.0 Å for reaction 1 as a common state of zero energy. Following the notation developed in Section 2.2, we refer to the PM3-SRP method optimized at the end of the FM stage in the *i*th RP-FM iteration as PM3-*i*, which is in turn used to update the reaction path in iteration *i*+1; the original PM3 method is referred to as PM3-0. For convenience, we also adopted a double-slash notation, where "A // B" is used to denote a potential energy profile computed at level A based on the reaction-path configurations determined at level B. In Figure 2a, we label the potential energy profiles computed with the PM3-*i* methods based on the reaction paths determined at a consistent level (i.e., PM3-*i* // PM3-*i*) as "PM3-*i*", whereas the SPE profiles computed at the HF/3-21G level for the same reaction-path configurations (i.e., HF/3-21G // PM3-*i*) are labeled as "HF-*i*". Note that the "PM3-*i*" reaction path is the one updated in the RP stage of the (*i*+1)th RP-FM iteration (see Figure 1). The potential energies of the stationary points as well as the locations of the reactant complex determined from the PM3-0, PM3-5, and HF/3-21G calculations are also given in Table 1 for comparison. The numerical values of the 21 PM3 parameters obtained in each RP-FM iteration are available in Table S2 of the Supporting Information.

Figure 2a shows that the reaction barrier height given by the PM3-0 method is 9.5 kcal/mol (Table 1), which is substantially higher than the barrier of 0.2 kcal/mol obtained from the corresponding HF/3-21G SPE profile (HF-0). After the first RP-FM iteration, the updated PM3 energy profile (PM3-1) is brought into a better agreement with the associated HF/3-21G SPE profile (HF-1). Such agreements are steadily improved over the RP-FM iterations. As shown in Figure 2a, after five RP-FM iterations, the energy profile from the optimized PM3-SRP (PM3-5) and that from the target HF/3-21G level (HF-5) are almost identical. The barrier height given by the PM3-5 method is 0.5 kcal/mol, which closely reproduces the target HF/3-21G barrier of 0.2 kcal/mol (Table 1).

The RP-FM optimization also brings the tailored SE methods to agree with target AI method on the barrier shape. As shown in Figure 2a, after five iterations, RP-FM successfully shifts the position of the reactant minimum from RC = −0.7 Å on the PM3-0 profile to RC = −0.4 Å on the PM3-5 profile, which is in an excellent agreement with the reactant position of RC = −0.3 Å determined at the target HF/3-21G level (Table 1).

As a result of RP-FM, both the barrier height and barrier shape from the HF/3-21G energy profile are reproduced by the optimized PM3-SRP methods. The energy profiles do not change significantly between consecutive iterations from PM3-4 to PM3-5, suggesting that a convergence is established.

The capability of RP-FM in refining multiple aspects of the reaction energy profiles is not a surprise, considering that a significantly amount of AI information (243 force components in the gas-phase PR-FM case) is utilized in parametrizing the SE-SRP methods. Moreover, we attribute the success to the iterative fashion in which RP-FM is conducted. In a qualitative sense, RP-FM iterations allow new information about the AI PES, "learned" by the SE-SRP method through the force-matching processes, to be used to locate a "better" reaction path, thereby upgrading the SE-SRP path to the AI quality.

The necessity of carrying out the RP-FM cycle iteratively can be further illustrated based on a hypothetical two-dimensional AI PES as depicted in Figure 3. In RP-FM, before convergence is established, the reaction path determined at an intermediate SE-SRP level is not an MEP on the target AI PES. Let the green dotted curve in Figure 3 represent the SE-0 path (see Section 2.2 for notation), which differs from the true AI MEP (shown as a yellow solid curve). In the first RP-FM iteration, calibrating SRPs against AI forces along the SE-0 path leads to the SE-1 method. Although the single RP-FM cycle make the SE-1 method to agree with the AI method along the SE-0 path, the potential energy profile associated with this path (i.e., SE-1 // SE-0) cannot reproduce the AI energy profile along the true AI MEP. As the SE-0 path (green) is not an MEP (yellow) on the AI PES, the "SE-1 // SE-0" energy profile would overestimate the target barrier height and misrepresent the target barrier shape. The quality of the SE-0 reaction path, however, can be improved by updating the path using the SE-1 method (during the RP stage of the second RP-FM iteration). Because the SE-1 method reproduces the AI forces associated with the SE-0 path (one exemplary force component is illustrated as the green arrow in Figure 3), under the influence of such forces the SE-0 path (now a high-energy path on the SE-1 surface) would slide down toward the valley region, thereby moving the recomputed path (i.e., the SE-1 path, shown as the red dashed curve) toward the true AI MEP (yellow). By conducting RP-FM iteratively, the SE-SRP path is expected to gradually converge to the true MEP on the target AI surface.

An idealized scenario of converging SE-SRP-level reaction paths to the target AI-level MEP during RP-FM iterations can be described as follows. Imagine that the RP stage in the *i*th RP-FM iteration generates an SE-SRP path, labeled as the SE-(*i*-1) path (following the notation developed in Section 2.2).

Table 1. Reaction Coordinates (RC) of Reactant Minima and Associated Barrier Heights in Gas Phase and Solution for Reaction 1[a]

	gas phase		solution	
methods	RC	$\Delta E^{\ddagger}$	RC	$\Delta G^{\ddagger}$
HF/3-21G[b]	−0.3	0.2	−0.52	2.5
PM3-0[c]	−0.7	9.5	−0.68	12.3
PM3-5[d]	−0.4	0.5	–	–
PM3-5S[e]	–	–	−0.50	3.0

[a]Reaction coordinates (RC) are in Å. Energy barriers ($\Delta E^{\ddagger}$) and free energy barriers ($\Delta G^{\ddagger}$) are in kcal/mol. [b]Based on the energy or potential mean force (PMF) profiles determined from separate simulations at the HF/3-21G or HF/3-21G/MM level. [c]PM3-0 denotes the original PM3 method. [d]PM3-5 denotes the PM3-SRP method optimized at the end of the fifth gas-phase RP-FM iteration. [e]PM3-5S denotes the PM3-SRP method optimized at the end of the fifth solution-phase RP-FM iteration.

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

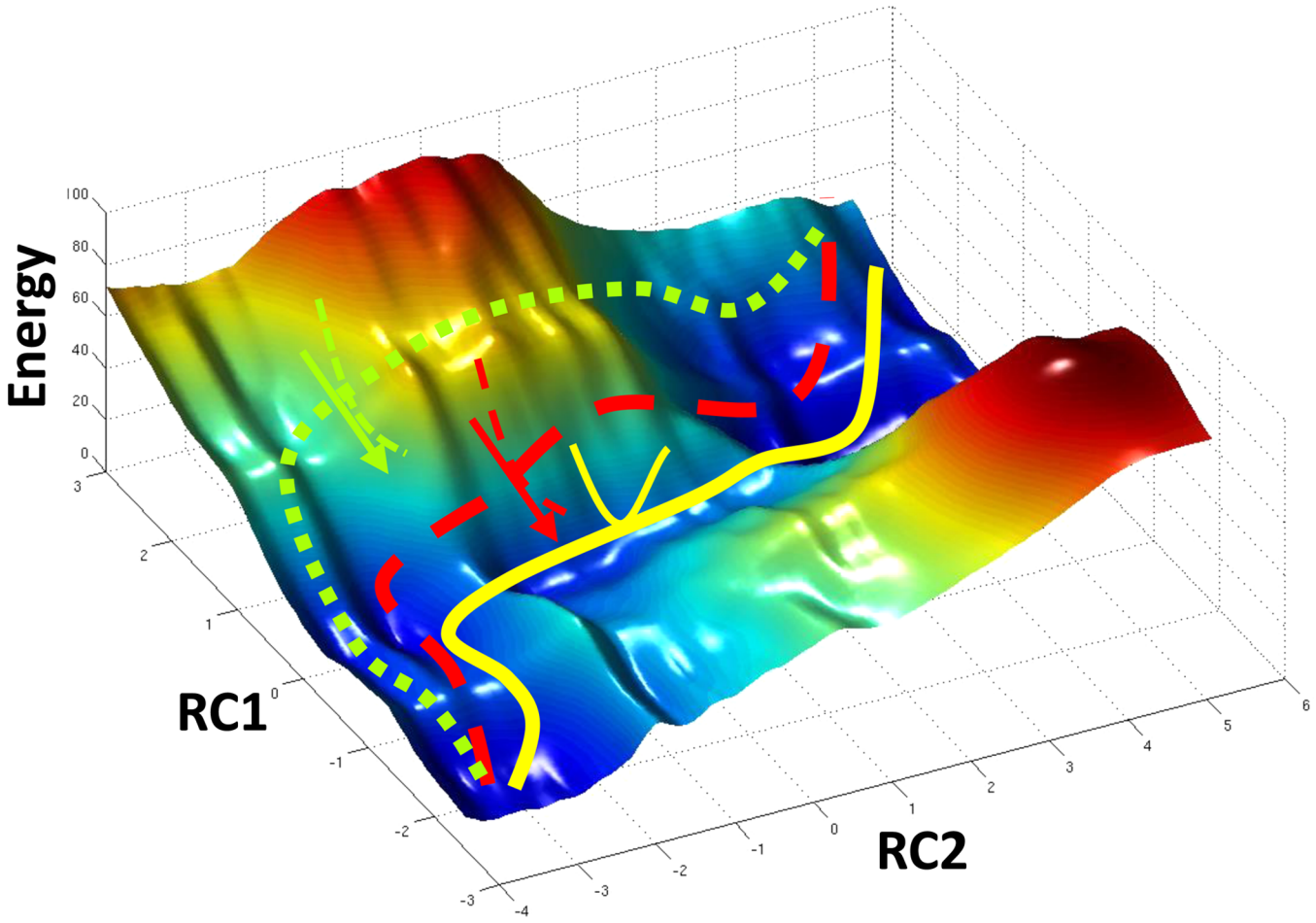


Figure 3. An illustration of the reaction path converging process over RP-FM iterations. The SE-0 path (green dotted curve), SE-1 path (red dashed curve), and true AI MEP (thick yellow solid curve) are shown on a hypothetical two-dimensional AI PES, for which the two reaction coordinates are labeled as RC1 and RC2, respectively. Calibrating SE-SRPs to reproduce AI forces along the SE-0 path leads to the SE-1 method. Although the single-point energy profiles "SE-1 // SE-0" and "AI // SE-0" agree with each other, they both overestimate the reaction barrier height associated with the true AI MEP. When the reaction path is recomputed with the SE-1 method, under the influence of the forces matched to their target AI values (a representative force vector is illustrated as the green arrow), the SE-0 path (green) is updated to the SE-1 path (red), thereby moving the SE-SRP path toward the true AI MEP (yellow). When RP-FM is conducted iteratively, SE-SRP paths gradually converge to the MEP on the target AI PES. Large perpendicular force deviations between the SE-SRP and the target AI levels are expected before the SE-SRP path converges to the true AI MEP. For the true AI MEP, forces along the MEP are everywhere tangent to the path, and all perpendicular force components vanish. Before convergence is achieved, the reaction path determined at the SE-1 level is not an MEP on the AI PES; therefore, single-point forces computed along the SE-1 path at the target AI level (AI // SE-1) would display large nonzero forces in directions perpendicular to the reaction path (a non-vanishing force vector is illustrated as the red arrow). The RP-FM iterations minimize force deviations both along and perpendicular to the reaction path, effectively moving the SE-SRP path toward the AI one. Upon convergence, the SE-SRP path closely resembles the true AI MEP, faithfully reproducing both types of force components at the target level.

Assuming the SE-(i-1) path is essentially identical to the MEP determined on the target AI surface, next we show further RP-FM iterations will not move the path anymore. The FM stage that immediately follows the RP stage in the ith iteration will result in the SE-i method, which is expected to reproduce the AI atomic forces along the SE-(i-1) path (therefore, the SE-i method also reproduces the AI atomic forces along the true AI MEP). Given the fact that an MEP is uniquely determined if the atomic forces for all the path configurations are known, the recomputed path in the RP stage of the (i+1)th RP-FM iteration, i.e., the SE-i path, will be identical to the AI MEP because the forces associated with the two paths have already been faithfully matched with each other in the previous iteration. As an immediate consequence, the SE-i and SE-(i-1) paths become identical, and further iterations of RP-FM will no longer change the path, and convergence of the SE-SRP paths to the target AI MEP is achieved.

5.1.2. Average Force Deviations. The convergence of the RP-FM procedure was also monitored by the force deviations between the PM3-SRP and HF/3-21G methods over iterations. In particular, we computed the force deviations between the tailored SE and target AI methods before force matching was conducted in each RP-FM cycle. We refer to these force deviations as the initial force deviations. Note that the reaction paths based on which we evaluated each set of initial force deviations were determined consistently on the corresponding PM3-SRP PES. For example, the initial force deviations for the ith RP-FM iteration were computed from PM3-i // PM3-i and HF/3-21G // PM3-i SPEF calculations (see Sections 2.2 and 5.1.1 for notations). Figure 2b displays the initial force deviations per Cartesian coordinate over five RP-FM iterations;

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

note that each of these deviations is numerically equal to the negative value of the fitness function χ defined in eq 3. Before the first iteration, the PM3-0 method yields a force deviation of 12 kcal/mol/Å averaged over nine configurations (243 Cartesians) along the MEP. After a single cycle of RP-FM, the initial force deviation is greatly reduced to 1.5 kcal/mol/Å. The initial force deviations stabilize at <1 kcal/mol/Å over the remaining RP-FM iterations, implying that the optimized SE-SRP methods generate steady performance across iterations in reproducing the target AI forces, which is another sign of convergence of the procedure.

5.1.3. Force Deviations on Individual Configurations. Decompositions of the initial force deviations to contributions from individual configurations are shown in Figure 2c. It is shown that the force deviations do not distribute uniformly along the MEP. For the PM3-0 method, the transition state (TS) configuration, located at RC = 0.0 Å, is associated with a force deviation of 11.3 kcal/mol/Å per Cartesian coordinate, which is slightly higher than that of 10.3 kcal/mol/Å for the reactant state configuration located at RC = −0.7 Å but lower than a maximal force deviation of 14.0 kcal/mol/Å found at RC = −0.3 Å for a nonstationary structure between the reactant state and TS. As shown in Figure 2c, with a few RP-FM iterations, remarkable agreements between the PM3-SRP and HF/3-21G forces were obtained not only for the two stationary species (reactant and TS) but also for all nonstationary configurations connecting them. As a result, the initial force deviations on all nine configurations along the whole MEP were reduced to only 1.5 kcal/mol/Å or less in the fifth iteration of the gas-phase RP-FM.

5.1.4. Force Deviations along and Perpendicular to the Reaction Path. Although force matching was conducted in Cartesian coordinates in the present study, other coordinate systems can be used. For example, with normal coordinates[57] defined for a given reaction path Hamiltonian,[77] the total atomic forces can be conceptually separated into two types of components: forces along and perpendicular to the reaction path. For a true MEP, forces are everywhere tangent to the path. Therefore, nonzero forces only exist in the direction parallel to the MEP, whereas force components orthogonal to the path all vanish.

As we discussed in Section 5.1.1, before convergence is established in RP-FM iterations, the reaction path determined on an SE-SRP surface is generally not an MEP on the target AI PES. Thus, the SE-SRP and AI methods are expected to differ in both types of force components. The source of the force deviations perpendicular to the reaction path is illustrated in Figure 3. When atomic forces are evaluated at the AI level along an intermediate SE-SRP path (e.g., the SE-1 path, shown as the red dashed curve in Figure 3), there would be large force deviations between the tailored and the target levels in the directions orthogonal to the path. Although such forces vanish on the corresponding SE-SRP PES (by definition of an MEP), they do not vanish on the AI PES (one of the nonvanishing force vectors perpendicular to the SE-1 path is depicted as a red arrow in Figure 3). The greater the perpendicular force deviations are, the more likely the SE-SRP path represents a high-energy path on the AI PES. When RP-FM converges the SE-SRP path to a true AI MEP (i.e., the "yellow" path in Figure 3), such deviations become zeroes because the perpendicular force components now vanish on both surfaces. On the other hand, integration of the parallel force components along the reaction coordinate leads to the energy profile; thus, the deviations of forces parallel to the reaction path correlate more with the difference in the energy profiles determined at the two levels. Altogether, force deviations of both types provide a combined measure of how well an SE-SRP reaction path resembles the MEP obtained on the target AI PES.

To understand whether the RP-FM approach is effective in matching both these types of force components, thereby converging an SE-SRP path to the target AI MEP, we estimated the force deviations along and perpendicular to the reaction path (SI.2 and Figure S1, Supporting Information). We found that initially the deviation on the perpendicular force components between the PM3-0 and HF-0 methods is significant (12.4 kcal/mol/Å) but still smaller than the deviation of the forces parallel to the reaction path (16.7 kcal/mol/Å). With five iterations of RP-FM, the PM3-5 method reproduces the target HF/3-21G forces so well that the deviations on both types of force components are greatly reduced to less than 1 kcal/mol/Å. These results indicate that the reaction path determined at the PM3-5 level closely mimics the one determined on the HF/3-21G PES.

5.1.5. Force Matching on Selected Configurations. A striking observation from the results presented in Section 5.1.1 is that although the whole RP-FM procedure uses no energy information as explicit input (i.e., SRPs are exclusively optimized based on atomic forces), the optimized SE-SRP methods faithfully reproduce the target energy profiles. This success is worth of some further analyses. For a reactive system that involves transitions among multiple stationary states (such as stable basins and transition states), force matching for stationary species alone is insufficient to uniquely determine the relative energies (or the density of states in a broader sense) among these species. This problem can be illustrated in Figure 4, which shows three schematic one-dimensional energy profiles that drastically differ in shape but still agree with one another in terms of forces at the stationary points. Intriguingly, RP-FM does not seem to suffer from this nonuniqueness problem when force matching is conducted for a chemical reaction, for which the involvement of multiple stationary states is unavoidable. This observation made us to hypothesize that RP-FM acquires its capability to "synthesize" the semiglobal features of the target PES through the special way force matching is handled. In the FM stage, many pieces of local force information based on interconnected configurations along a reaction path are combined together such that the density of states of the involved species (stationary as well as nonstationary) is faithfully recovered. If this hypothesis is valid, RP-FM is expected to become less effective when force matching is conducted only on "isolated" structures that do not overlap in the configuration space.

To test the above hypothesis, we performed four additional RP-FM calculations as controls, where we only included the reactant state (R), transition state (TS), both reactant and transition state (R+TS), or reactant and transition state plus an additional point along the MEP (R+TS+1p) in the FM stages. The potential energy profiles obtained by using these selected configuration RP-FM schemes are displayed in Figure 5. Our results show that neither the R-only (Figure 5a) nor the TS-only (Figure 5b) scheme can reproduce the target HF/3-21G energy profiles. While having both stationary species included (R+TS) improves the results (Figure 5c), only when an extra point between them is added for force matching (R+TS+1p) does stable convergence behavior start to emerge (Figure 5d). These observations strongly support the validity of the

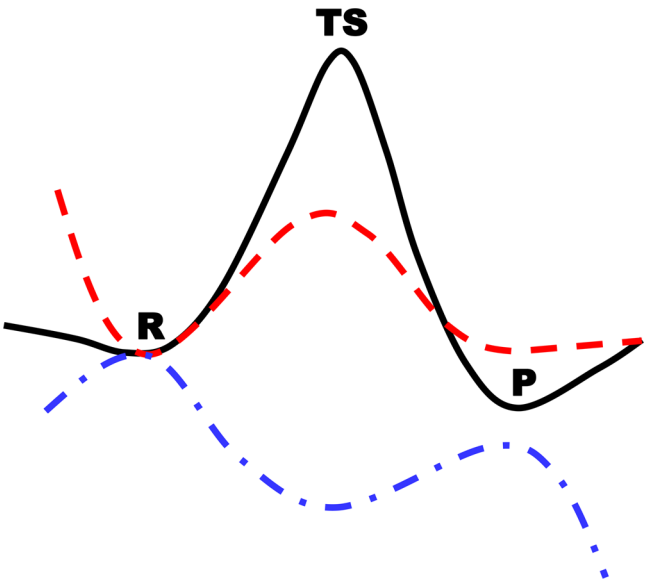


Figure 4. A schematic illustration of the challenge in "reactive" force matching. For a simple reaction process involving a reactant (R), a product (P), and a transition state (TS), although the two possible SE-SRP methods (the associated energy profiles are shown as the red dashed and blue dot-dashed curves) match the atomic forces at the target AI level for all three stationary points (R, TS, and P), they both fail to reproduce the overall shape and quantitative features of the target AI energy profile (the black solid curve). Thus, force matching based on non-overlapping regions in the configuration space is expected to generate arbitrary energy profiles. By contrast, by combining many pieces of local force information obtained for interconnected configurations along a continuous reaction path, RP-FM is capable of building the semiglobal features of the target AI PES such that the relative energies (density of states) for multiple stationary and nonstationary species at the target level are also faithfully reproduced.

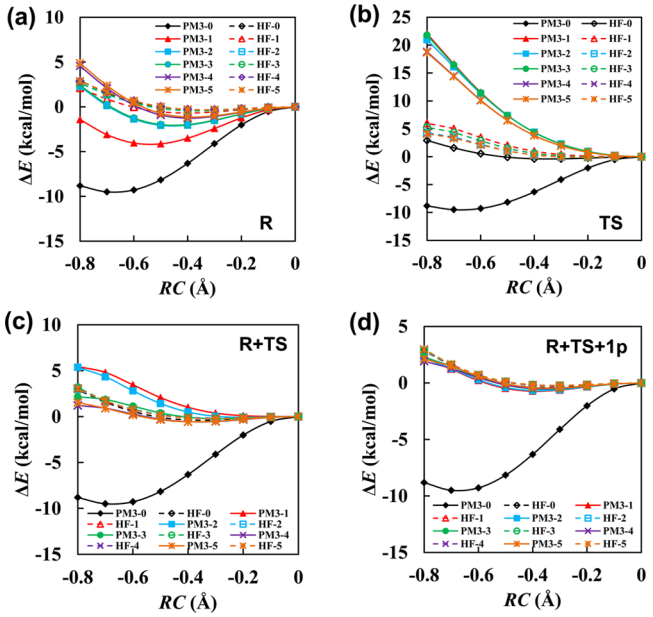


Figure 5. Potential energy profiles of reaction 1 over gas-phase RP-FM iterations by using selected-configuration force-matching schemes, in which only (a) the reactant state (R), (b) transition state (TS), (c) reactant and transition state (R+TS), or (d) reactant and transition state plus an additional point (R+TS+1p) are included in the FM stage of each RP-FM cycle in optimizing PM3-SRPs. Neither the R-only nor the TS-only scheme can reproduce the target HF/3-21G energy profiles. While having both stationary species included (R+TS) improves the results, only when an extra point between them is added for force matching (R+TS+1p) does stable convergence behavior start to emerge.

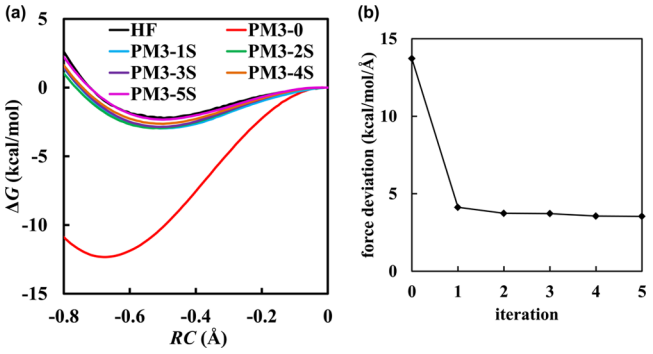


Figure 6. (a) Potential of mean force (PMF) profiles for reaction 1 obtained from QM/MM umbrella sampling simulations based on the original PM3 method (PM3-0), PM3-SRP methods optimized from *i*th (i = 1–5) solution-phase RP-FM iteration (PM3-*i*S), and HF/3-21G method (HF). (b) Initial force deviations (averaged based on 1,890 force components on 70 configurations along each reaction path) over five solution-phase RP-FM iterations.

hypothesis we formulated. Force matching on a multiple-state system is most effective when it is conducted on an ensemble of "overlapping" configurations. On the basis of this "maximal overlap" principle, our design of conducting force matching along a continuous reaction path in RP-FM is nearly optimal. Additional analyses of force deviations for these selected configuration RP-FM calculations can be found in SI.3 and Figures S2–S5 of the Supporting Information.

5.2. RP-FM in Solution. Next, we present the results of applying RP-FM on a condensed-phase chemical reaction by making use of combined QM/MM simulations; this corresponds to the "finite-temperature" version of the procedure that aims at reproducing the potential of mean force (PMF) profiles at the target AI/MM level. In the condensed-phase RP-FM, force matching is directly conducted on configurations sampled along a free energy path with SE/MM simulations. In particular, we employed RP-FM to optimize PM3-SRPs against HF/3-21G/MM forces for reaction 1 in aqueous solution.

5.2.1. Potential of Mean Force Profiles. The PMFs based on the PM3-SRP methods obtained from five iterations of solution-phase RP-FM are displayed in Figure 6a. Based on the notation in Section 2.2, we refer to the PM3-SRP method optimized at the end of the *i*th RP-FM iteration as PM3-*i*S, where the suffix "S" is added here to distinguish the solution-phase-optimized SRPs from those obtained in the gas phase. In Figure 6a, the PMFs determined by employing the original PM3, PM3-*i*S, and HF/3-21G methods in QM/MM simulations are labeled as PM3-0, PM3-*i*S, and HF, respectively. It is shown that initially the PM3-0 PMF differs substantially from the HF one. The RP-FM optimization dramatically improves the agreement between the PMF profiles obtained at the PM3-SRP/MM and HF/3-21G/MM levels. At the end of the fifth RP-FM iteration, the PMF produced from the PM3-5S method is essentially identical to that separately determined at the target HF/3-21G/MM level.

Locations of the reactant free energy minima as well as the free energy barrier heights obtained from the PM3-0, PM3-5S,

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038–3054

and HF/3-21G methods are also compared in Table 1. The free energy of activation ($\Delta G^{\ddagger}$) for reaction 1 determined at the HF/3-21G/MM level is 2.5 kcal/mol, slightly lower than the value of 3.3 kcal/mol reported by Mo and Gao.[61] This difference could be caused by the use of a smaller sample size of 30,000 configurations (30 ps) per window in our HF/3-21G/MM umbrella sampling MD simulations compared with a sample size of 2×10^6 configurations collected per window in Mo and Gao's Monte Carlo simulations. In contrast to the HF/3-21G method, the PM3-0 method significantly overestimates the PMF barrier by 9.8 kcal/mol, giving rise to a free energy of activation as high as 12.3 kcal/mol. After a few RP-FM iterations, the optimized PM3-SRP methods faithfully reproduce the target PMF barrier height. For example, the free energy barriers obtained by using the PM3-4S and PM3-5S methods are 2.9 and 3.0 kcal/mol, respectively, which may only differ from the benchmark HF/3-21G results within the margin of statistical error (Figure 6a).

In addition to the improved free energy barrier heights, Figure 6a and Table 1 also show the RP-FM method's capability in shifting the position of the reactant minimum toward its AI/MM value, therefore precisely reproducing the shape of the target free energy profile. On the basis of the benchmark PMF obtained at the HF/3-21/MM level, we located a reactant complex at RC = −0.52 Å, which is in agreement with the literature value reported in Mo and Gao's work.[61] This is in marked contrast with the result from the PM3-0 method, which predicts a significantly "earlier" reactant free energy minimum at RC = −0.68 Å. After five iterations of RP-FM, the PM3-5S method successfully moves the reactant minimum to RC = −0.50 Å on the corresponding PMF, in excellent agreement with the benchmark PMF result (Table 1). Again, we attribute the success of the optimized SE-SRP method in reproducing this positional feature of the target PMF to the iterative nature of the RP-FM process.

Figure 6b shows the initial force deviations between the tailored and target methods over solution-phase RP-FM iterations; these force deviations were averaged over 70 configurations (1,890 Cartesian components) sampled along the reaction path in solution. Before any SRP optimization, the PM3-0 method displays a large initial force deviation of 14 kcal/mol/Å, which is then reduced to and stabilized at about 4 kcal/mol/Å during RP-FM. Due to the inclusion of a larger number of configurations in force matching, the solution-phase RP-FM seems to converge at a somewhat greater force-deviation value, whereas the force deviations in the gas-phase RP-FM stabilize at <1 kcal/mol/Å (Figure 2b).

Taken together, the agreement between the PMF profiles obtained from HF/3-21G/MM and from the PM3-5S/MM simulations as well as the greatly reduced force deviations between the two methods indicate that the employment of the RP-FM strategy in solution is successful. The numerical values of the 21 PM3 parameters obtained in each solution-phase RP-FM iteration can be found in Table S3 of the Supporting Information.

5.2.2. Mulliken Charges. The ensemble-averaged Mulliken charges on the NH_4^+ and NH_3 groups from QM/MM simulations based on the original PM3 (PM3-0) and the solution-phase optimized PM3-SRP (PM3-5S) methods are compared with the results from the corresponding HF/3-21G/MM SPE calculations (HF-0 and HF-5) in Figure 7. The standard deviations on these group charges are generally small, ranging from 0.01 to 0.04 *e* and from 0.01 to 0.03 *e* in the PM3(-SRP)/MM and HF/3-21G/MM results, thus are not shown in the figure.

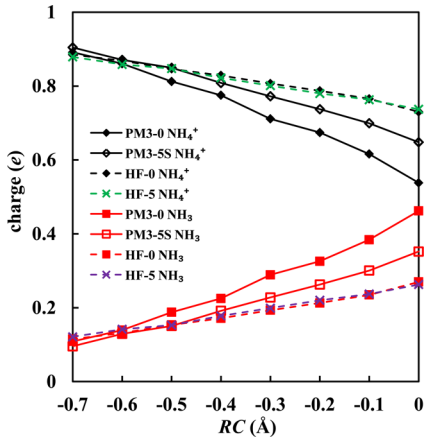


Figure 7. Average Mulliken charges on the NH_4^+ and NH_3 groups along the proton transfer reaction coordinate RC, obtained from QM/MM simulations of reaction 1 in solution by using the original PM3 method (PM3-0) and the solution-phase RP-FM optimized PM3-SRP method (PM3-5S), in comparison with the HF/3-21G benchmark results. The benchmark charges, labeled as HF-0 and HF-5, were obtained from single-point HF/3-21G/MM calculations based on the configurations collected from the PM3-0/MM and PM3-5S/MM simulations, respectively.

The PM3-0 method predicts a smaller charge difference between the two groups compared with the HF/3-21G results. This underestimated charge difference becomes more pronounced when the system evolves from the reactant to the transition state, leading to a maximal charge deviation between the two methods at RC = 0.0 Å. For example, at RC = −0.50 Å, the PM3-0/MM simulations result in an average charge of 0.82 and 0.18 *e* on the NH_4^+ and NH_3 group, respectively (Δq = 0.64 *e*), compared with the corresponding charges of 0.85 and 0.15 *e* (Δq = 0.70 *e*) at the HF/3-21G/MM level. At the transition state (RC = 0.0 Å), the PM3-0/MM simulations yield an average charge of 0.52 and 0.48 *e* on the NH_4^+ and NH_3 group, respectively (Δq = 0.04 *e*), implying that the transferred proton only bears a small positive charge, whereas the HF/3-21G/MM calculations produce a charge of 0.78 and 0.22 *e* on the two groups (Δq = 0.56 *e*). These results suggest that the PM3-0 method underestimates the charge separation of the two groups when proton transfer occurs.

With five iterations of RP-FM optimizations, the agreement between the PM3-5 and HF/3-21G methods on the group charge distributions is significantly improved. It is shown in Figure 7 that the Mulliken charge curves produced by the PM3-5S/MM simulations resemble the HF/3-21G/MM results to a greater extent than by using the PM3-0/MM simulations. Consequently, the PM3-5S method predicts an average charge of 0.65 and 0.35 *e* on the NH_4^+ and NH_3 group, respectively (Δq = 0.30 *e*), at the transition state; the improved charge distributions are expected to offer an electrostatic description upgraded toward the benchmark level.

5.2.3. RP-FM Improves both Gas-Phase and Solvation Barriers. An important issue remaining to be addressed is whether the RP-FM optimized SE-SRP methods can be used to describe the polarization effects arising from the solvent (or protein) environment in condensed phases. Reliable descriptions of such effects are of fundamental importance for

dx.doi.org/10.1021/ct4009624 | J. Chem. Theory Comput. 2014, 10, 3038−3054

Table 2. Gas-Phase Solute and Solvation Energies of Reactant and Transition State, with Their Contributions to Free Energy Barriers Estimated from QM/MM Simulations Using Original PM3 (PM3-0), Solution-Phase RP-FM-Optimized PM3-SRP (PM3-5S), and Target HF/3-21G Methods[a]

	PM3-0			HF/3-21G		
	$\langle E_{gas}\rangle$[b]	$\langle E_{SX}\rangle$	$\langle E_{tot}\rangle$	$\langle E_{gas}\rangle$	$\langle E_{SX}\rangle$	$\langle E_{tot}\rangle$
reactant (RC_R = −0.68 Å)[c]	138.4 [1.7]	−110.6 [7.8]	31.6 [8.3]	−70374.8 [2.9]	−127.8 [8.8]	−70502.6 [9.2]
reactant (RC_R = −0.50 Å)[d]	139.8 [1.5]	−103.2 [8.0]	34.7 [8.1]	−70378.0 [2.4]	−125.9 [9.5]	−70503.9 [9.8]
TS[e]	146.8 [1.3]	−97.8 [7.6]	50.0 [7.9]	−70377.1 [2.2]	−120.6 [8.9]	−70497.9 [9.4]
$\Delta g^{\ddagger}$ (RC_R = −0.68 Å)[f]	**8.4**	**12.8**	**18.4**	−2.3	7.2	4.7
$\Delta g^{\ddagger}$ (RC_R = −0.50 Å)[g]	7.0	5.4	15.3	**0.9**	**5.3**	**6.0**
	PM3-5S			HF/3-21G		
	$\langle E_{gas}\rangle$	$\langle E_{SX}\rangle$	$\langle E_{tot}\rangle$	$\langle E_{gas}\rangle$	$\langle E_{SX}\rangle$	$\langle E_{tot}\rangle$
reactant (RC_R = −0.50 Å)[h]	−43.6 [1.6]	−104.0 [7.8]	−148.1 [1.1]	−70378.4 [2.1]	−127.2 [9.1]	−70505.7 [9.2]
TS	−43.4 [1.3]	−101.4 [7.3]	−143.6 [8.2]	−70377.9 [2.3]	−125.7 [9.3]	−70503.5 [9.4]
$\Delta g^{\ddagger}$ (RC_R = −0.50 Å)	**0.2**	**2.6**	**4.5**	**0.5**	**1.5**	**2.2**

[a]Energies are in kcal/mol; standard deviations of average energies are given in the square brackets; the recommended data for comparing the performance of the PM3-SRP and HF/3-21G methods are highlighted in bold. [b]⟨...⟩ denotes an ensemble average. [c]Reactant (R) configurations were sampled at the PM3-0/MM level about RC_R = −0.68 Å (the reactant position located on the PM3-0 PMF). [d]Reactant configurations were sampled at the PM3-0/MM level about RC_R = −0.50 Å (the reactant position located on the HF PMF). [e]Transition state (TS) configurations were always sampled about RC = 0.0 Å for the symmetric proton transfer in reaction 1. [f]Gas-phase solute and solvation contributions to the free energy barriers were estimated by using the reactant ensemble at RC_R = −0.68 Å. [g]Gas-phase solute and solvation contributions to the free energy barriers were estimated by using the reactant ensemble at RC_R = −0.50 Å. [h]Reactant configurations were sampled at the PM3-5S/MM level about RC_R = −0.50 Å (the reactant positions located consistently on both the PM3-5S and HF PMFs)

understanding enzyme catalysis or solvation phenomena in general.

In the present study, because only the total force experienced on each solute atom is matched between the SE-SRP/MM and the target AI/MM levels, it is not known *a priori* that the force components associated with the gas-phase and solvation contributions are both fit in a balanced manner. A minimally acceptable outcome occurs when the RP-FM procedure does not significantly alter the relative contributions of the gas-phase and solvation terms to the free energy barriers such that similar solvation effects are obtained from the resulting SE-SRP and the original SE methods. It is even better if the RP-FM optimizations upgrade the solvation description to the target AI level. On the contrary, if the balance of these contributions is destroyed or compromised to a significant extent, the resulting SE-SRPs would become of little use in providing the correct physics and should be discarded, no matter how well they reproduce the total forces and the PMF profiles at the target level.

It is important, however, to point out that the above issue is *not* an inherent problem of RP-FM. The RP-FM strategy by itself is very general and therefore can be implemented with great flexibility. For example, an alternative force-matching scheme can be adopted to remove the above concern. Because the total atomic force can be readily decomposed to various components, such as those arising from the gas-phase and solvation contributions, one can fit these force components separately to their target AI/MM values in RP-FM. Such a component-based force-matching scheme is desirable to maintain a balanced improvement on forces based on their physical origins. In the present study, for simplicity, we only matched the total forces, instead of the decomposed ones, between the tailored and target levels. To verify that this treatment does not generate spurious solvation effects to the free energy profiles, we examined the gas-phase solute and solvation barriers separately for the related methods.

To dissect these effects, we estimated the gas-phase and solvation contributions to the free energy barriers with the decomposed barrier components $\Delta g^{\ddagger}_{gas}$ and $\Delta g^{\ddagger}_{SX}$ defined in eqs 7 and 8 (Section 4.9). The gas-phase barrier $\Delta g^{\ddagger}_{gas}$ (eq 7) keeps track of the solute energy E_{gas} (eq 6) change in the absence of solvent polarizations when the system evolves from the reactant (R) to the transition state (TS) ensemble, and the solvation barrier $\Delta g^{\ddagger}_{SX}$ (eq 8) records such changes based on the total QM/MM solute−solvent interaction energy E_{SX} (eq 4).[78] The related results for the original PM3 (PM3-0), PM3-5S, and HF/3-21G methods are compared in Table 2 to monitor the details of how the PM3 method is improved before and after the solution-phase RP-FM optimizations. Note that in this work, we only focus on the gas-phase and overall solvation contributions; the more detailed QM/MM energy decomposition that breaks the aqueous polarization effects into various individual terms[76] is left for future studies.

The PM3-0 method locates a reactant free energy minimum at RC = −0.68 Å and gives rise to a PMF barrier height of 12.5 kcal/mol (Figure 6a and Table 1), for which the estimated gas-phase solute contribution $\Delta g^{\ddagger}_{gas}$ to the free energy barrier is 8.4 kcal/mol. On the basis of the same reactant and transition state ensembles, the PM3-0 method yields a solvation barrier $\Delta g^{\ddagger}_{SX}$ of 12.8 kcal/mol, indicating that the reactant state is better solvated than the TS and thus solvation further elevates the free energy barrier. By contrast, the corresponding contributions computed at the target HF/3-21G/MM level, if the position of the reactant minimum RC = −0.68 Å as determined on the PM3-0 PMF is used, yield an inverse gas-phase barrier ($\Delta g^{\ddagger}_{gas}$ = −2.3 kcal/mol, i.e., the gas-phase solute is less stable in the reactant state than in the TS) and a significant smaller solvation barrier ($\Delta g^{\ddagger}_{SX}$ = 7.2 kcal/mol). The data imply that the PM3-0 method predicts a stronger desolvation effect upon formation of the TS compared with the HF/3-21G benchmark. Taken together, these results suggest that both the gas-phase and solvation terms contribute to the large discrepancy between the PM3-0/MM and HF/3-21G/MM PMF barriers.

We note that the shape of the PMF produced by PM3-0 differs significantly from that separately determined at the HF/3-21G level, which locates a reactant free energy minimum at a

RC value of −0.50 Å. Therefore, the above discrepancy between the PM3-0 and HF/3-21G barriers is coupled with the fact that the reactant state identified on the PM3-0 PMF is not a stationary species on the HF/3-21G/MM surface. In this case, the reactant configurations found on the PM3-0/MM PES (RC = −0.68 Å) correspond to high-energy structures on the HF/3-21/MM surface; this explains the inverse gas-phase barrier contribution of −2.3 kcal/mol at the HF/3-21G/MM level when the positional variation effect of the reactant state is not taken into account.

To explore other possibilities of comparing the two methods with less ambiguity, we also computed the decomposed barrier contributions based on the reactant position of RC = −0.50 Å identified on the HF PMF. This choice of the reactant ensemble yields a gas-phase barrier ($\Delta g^{\ddagger}_{gas}$) of 7.0 kcal/mol for the PM3-0 method, which is similar to the value determined with RC = −0.68 Å (8.4 kcal/mol) but still significantly higher than the gas-phase barrier of 0.9 kcal/mol benchmarked at the HF/3-21G/MM level. With the choice of RC = −0.50 Å, the solvation barrier ($\Delta g^{\ddagger}_{SX}$) produced by the PM3-0 method is now reduced to 5.4 kcal/mol, significantly lower than that obtained by using RC = −0.68 Å (12.8 kcal/mol); correspondingly, the solute/solvent interactions determined from the HF/3-21G/MM calculations yield a solvation barrier of 5.3 kcal/mol. When RC = −0.50 Å is used, it appears that the deviation between the PM3-0 and HF free energy barriers is dominated by the gas-phase contribution because the solvation contributions from the two methods now agree better with each other; such agreement, however, is rather fortuitous, considering that the configurations found at RC = −0.50 Å also do not represent a stable reactant species on the PM3-0 PMF.

Based on the above analyses, we decided that the most consistent comparison of the free energy barrier components between the PM3-0 and HF/3-21G methods should be made with the choice of reactant configurations at RC = −0.68 Å for PM3-0 and at RC = −0.50 Å for HF/3-21G, respectively, and with the choice of TS configurations at RC = 0.0 Å for both methods. The related data in Table 2 are shown in bold. From these highlighted results, we can see that the PM3-0 method overestimates the free energy barrier because of two factors; both the gas-phase and solvation barriers are overestimated at the PM3-0/MM level by 7.5 kcal/mol compared with the HF/3-21G/MM benchmark values. The combined error from these two contributions in PM3-0, estimated by $\Delta g^{\ddagger}_{tot}$ (eq 9), results in a free energy barrier of 18.4 kcal/mol, which is 12.4 kcal/mol higher than the same estimate obtained at the HF/3-21G/MM level (6.0 kcal/mol) (Table 2), satisfactorily explaining the PMF barrier difference of 9.8 kcal/mol (i.e., 12.3 vs 2.5 kcal/mol; Table 1) between the two methods.

Next, we show how RP-FM improves the performance of the optimized PM3-SRP method by reducing both errors associated with the gas-phase solute and solvation barriers. For the PM3-5S method optimized by the solution-phase RP-FM, because the shape of free energy profile at the target HF/3-21G/MM level is faithfully reproduced, the position of the reactant state is located at RC = −0.50 Å on both the PM3-5S and HF PMFs; this consistency allows us to compare the barrier components without the ambiguity discussed above for PM3-0. With the PM3-5S method, the gas-phase solute energy contributes 0.2 kcal/mol to the overall barrier, in a close agreement with the gas-phase barrier contribution of 0.5 kcal/mol in the HF/3-21G benchmark (Table 2). The solvation barrier produced by the PM3-5S method is 2.6 kcal/mol, in an excellent agreement with the HF/3-21G result of 1.5 kcal/mol. In contrast to the PM3-0 method, which displays large errors of 7.5 kcal/mol on both the gas-phase and solvation barriers, the PM3-5S method simultaneously reduce both errors to about 1 kcal/mol. The combined error in the free energy barrier, estimated from $\Delta g^{\ddagger}_{tot}$, is reduced to 2.3 kcal/mol in PM3-5S, compared with the large error of 12.4 kcal/mol in PM3-0. Altogether, these results demonstrate that not only does the RP-FM optimized SE-SRP method reproduce the target AI PMF, the individual gas-phase and solvation contributions of the free energy barrier are also upgraded in a balanced way to reach their target-level accuracy. This finding is indeed very encouraging, especially considering that the associated force components were not fit separately in the present RP-FM scheme.

6. RELATION TO OTHER WORKS

Before we conclude, we discuss the differences between the present work and other related investigations in the literature. The force-matching technique has been discussed in the context of QM/MM calculations.[42,43,45] These existing studies, however, focused on using QM/MM as an efficient target potential to tailor classical MM force fields or empirical potentials.

For example, in the force-matching approaches of Rothlisberger and co-workers[43] and Wang and co-workers,[45] classical force field parameters are adjusted with an aim to simulate physical processes, such as bulk properties, binding affinities, etc.; the single-state nature limits these work to nonreactive processes that do not involve transitions among multiple energy/free energy basins. By contrast, our development targets simulating chemical processes of bond rearrangements, the proper description of which cannot be offered by classical MM potentials. The RP-FM method reported here, in spirit similar to pioneering work of Voth and co-workers,[47] matches forces between a tailored reactive potential and a reference QM potential. In the force-matched multi-state EVB (FM-MS-EVB) method developed by Voth and co-workers, the reactivity of the tailored potential is introduced through a reactive force field.[47] In our RP-FM approach, the tailored potential itself is also an electronic structure-based QM potential (i.e., SE), although obtained only at a fraction of the computational cost of the target QM potential. A "reactive" scheme is then employed in RP-FM so that force matching is conducted for the reactant, product, and transition state, as well as all nonstationary states along a reaction path that connects these regions. The combination of using QM/MM methods in both the tailored and reference potentials and force fitting along a reaction path makes RP-FM a distinctive approach that enables one to carry out "reactive" force matching.

There is a fundamental issue that needs to be addressed in the "reactive" force matching. We note that force matching seems to be a self-justified strategy for tuning force field parameters when it is applied to a non-reactive problem, as the dynamics of the system, restricted around a single basin, can be fully determined from the forces. Thus, no additional energy matching is needed. It is unclear, however, whether such a force-only matching scheme would still be sufficient when multiple stationary states are involved. In an extreme case, force matching for two non-overlapping states in the configuration space can be separately achieved to reproduce the local shape of the potential energy surfaces but fail to properly join their relative density of states (e.g., the problem illustrated in Figure

4). Our tests of only including selected structures along the reaction path in force matching, as presented in Section 5.1.5 and Figure 5, suggest that reproducing target forces is sufficient, provided that the structures for force matching overlap properly in the configuration space. In the RP-FM approach, force matching is conducted along a continuous reaction path; therefore the structural "overlap" is maximized.

The good performance of RP-FM in reproducing the target-level energy and free energy profiles in the gas phase and in solution also confirms that the force-only matching scheme is sufficient to restore the target-level density of states for a chemically reactive system, especially when force matching is used in conjunction with reaction paths. In the present RP-FM scheme, we only allow the use of the force information along a single reaction path when an SE-SRP is optimized. For example, in the *i*th RP-FM iteration, only the target AI forces determined along the SE-(*i*-1) path is used to tune the SE-*i* method. In principle, to maximize the utilization of target AI data, one may consider to include the target forces on all previous paths in optimizing the SE-SRP. Such a cumulative force-matching scheme, however, is under the risk of using non-overlapping structures if the SE-SRP paths in previous iterations significantly differ from one another. Therefore, the "cross-talk" among SE-SRP paths in force matching is not used in the present work.

The RP-FM method also differs from the "learn-on-the-fly" (LOTF) approach[42] in two aspects; the latter method has been demonstrated in simulating diffusion of point defects on solid surfaces. First, the tailored potentials in the LOTF approach are still empirical functions, which are made "reactive" by employing time-dependent parameters, whereas what we chose to parametrize is a true QM potential (i.e., an SE method) based on electronic wave functions variationally optimized from self-consistent field calculations. Second, the LOTF method relies on instantaneous force matching along short MD trajectories (a few time steps), which are constantly updated during MD simulations; such a force-matching scheme may lead to energy conservation problems in the resulting simulations due to the use of time-dependent Hamiltonians.

The iterative nature of fitting the force and resampling the potential energy surface in our work resembles the spirit of the optimal potential (OP) method[44] in solid phase simulations and that of the adaptive force-matching (AFM) method developed by Wang and co-workers.[45] A similar adaptive potential refining procedure has also been employed in the "paradynamics" method pioneered by Warshel and co-workers,[79] in which empirical valence bond (EVB) potentials in conjunction with Gaussian-type corrections are parametrized against QM/MM data in the free energy perturbation (FEP) framework.

7. CONCLUDING REMARKS

In this article, we have presented a novel approach, named Reaction Path Force Matching (RP-FM), with which semi-empirical QM (SE) methods can be reparameterized to reach the accuracy of *ab initio* QM (AI) methods for simulating specific chemical reactions efficiently and reliably. The RP-FM strategy enables one to sample the configuration space using an efficient SE method and subsequently tune the SE-SRP method to faithfully reproduce high-level AI forces along reaction pathways. Directly sampling the computationally expensive AI surfaces is thus circumvented. We have demonstrated the power of this approach in modeling the proton transfer reaction from ammonium to ammonia in the gas phase and in solution. The results suggest that the goal of obtaining energy or free energy profiles at an accuracy of AI(/MM) at the computational cost of SE(/MM) is satisfactorily achieved. The RP-FM approach is expected to have potential applications in simulating complex reactions in condensed phases, such as enzyme-mediated ATP hydrolysis,[80] which has posed a grand challenge to the traditional SRP fitting strategy. As the next step toward achieving this long-term goal, further tests of the RP-FM approach on several well-characterized enzymatic reactions are highly desirable.

Although the RP-FM approach is presented here for parametrizing semiempirical QM methods, the strategy of conducting force matching along a transition pathway can be possibly generalized to other multi-scale simulation frameworks that require synthesizing the efficiency and accuracy of more than one theoretical level. For example, instead of tuning an SE method, one can employ the RP-FM strategy to parametrize reactive force fields[81] against AI/MM data along reaction paths to gain further computational efficiency. Another potential application of the RP-FM strategy is to use it to calibrate coarse-grained (CG) potentials[41] against all-atom force fields for "conformationally reactive" processes. For simulating complex biological systems that undergo conformational changes, multi-basin CG potentials[82,83] can often be used to overcome the large time-scale limit associated with sampling the conformational transition pathways. When quantitative features of the conformational changes are of primary interest, the success of the CG model depends on how reliably it reproduces all-atom models in generating the transition pathways and the associated free energy profiles. By employing the RP-FM approach, one can obtain the CG representation of a conformational pathway, compute the all-atom forces along the pathway sampled at the CG level, map the all-atom forces to the CG particles, and tune the parameters in the CG model to match the forces coarse grained from the all-atom calculations. Such cycles can be repeated iteratively, analogous to the way RP-FM is carried out in fitting SE-SRPs against AI forces.

■ ASSOCIATED CONTENT

Supporting Information

List of PM3 parameters adjusted (Table S1). PM3-SRPs optimized along RP-FM (Tables S2 and S3). Decomposition of force deviations (Figure S1). Results for selected configuration RP-FM (Figures S2−S5). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jpu@iupui.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Wei Yang and Prof. Jiali Gao for helpful discussions. This work was supported by a start-up grant from Indiana University−Purdue University Indianapolis (IUPUI). The computing time was provided by the High-Performance Computing Cluster funded by the School of Science at IUPUI and by the BigRed High Performance Computing facilities at

Indiana University, funded by the National Science Foundation (NSF).

REFERENCES

(1) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645.
(2) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(3) Field, M. J.; Bash, P. A.; Karplus, M. J. *J. Comput. Chem.* **1990**, *11*, 700.
(4) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(5) *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J.; Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998.
(6) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
(7) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
(8) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley: New York, 1986.
(9) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.
(10) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1994.
(11) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
(12) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.
(13) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2002**, *117*, 8694.
(14) Thiel, W. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; Wiley: Chichester, 2003; Vol. *2*, p 487.
(15) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
(16) Bredow, T.; Jug, K. *Theor. Chem. Acc.* **2005**, *113*, 1.
(17) Thiel, W. *WIREs Comput. Mol. Sci.* **2014**, *4*, 145.
(18) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
(19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(20) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
(21) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601.
(22) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 817.
(23) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.
(24) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391.
(25) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616.
(26) Kolb, M.; Thiel, W. *Theor. Chem. Acc.* **1993**, *14*, 775.
(27) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495.
(28) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *Biochemistry* **2003**, *42*, 13558.
(29) Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2003**, *107*, 9567.
(30) Gonzalez-Lafont; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618.
(31) Nam, K.; Cui, Q.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486.
(32) Doron, D.; Major, D. T.; Kohen, A.; Thiel, W.; Wu, X. *J. Chem. Theory Comput.* **2011**, *7*, 3420.
(33) Wu, X.; Thiel, W.; Pezeshki, S.; Lin, H. *J. Chem. Theory Comput.* **2013**, *9*, 2672.
(34) Arabi, A. A.; Matta, C. F. *J. Phys. Chem. A* **2009**, *113*, 3360.
(35) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2005**, *123*, 164108.
(36) Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2011**, *7*, 857.
(37) Colvin, M. E.; Evleth, E.; Akacem, Y. *J. Am. Chem. Soc.* **1995**, *117*, 4357.
(38) Rudack, T.; Xia, F.; Schlitter, J.; Kötting, C.; Gerwert, K. *Proc. Natl. Acac. Sci. U.S.A.* **2012**, *109*, 15295.
(39) Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583.
(40) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.
(41) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
(42) Csányi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* **2004**, *93*, 175503.
(43) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 628.
(44) Laio, A.; Bernard, S.; Chiarotti, G. L.; Scandolo, S.; Tosatti, E. *Science* **2000**, *287*, 1027.
(45) Arkin-Ojo, O.; Song, Y.; Wang, F. *J. Chem. Phys.* **2008**, *129*, 064108.
(46) Doemer, M.; Maurer, P.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U. *J. Chem. Theory Comput.* **2014**, *10*, 412.
(47) Knight, C.; Maupin, C. M.; Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 3223.
(48) Truhlar, D. G.; Kuppermann, A. *J. Am. Chem. Soc.* **1971**, *93*, 1840.
(49) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363.
(50) Heidrich, D. In *The Reaction Path in Chemistry: Current Approaches and Perspectives*; Heidrich, D., Ed.; Kluwer: Dordrecht, 1995; p 1.
(51) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109*, 6688.
(52) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106.
(53) Hu, H.; Lu, Z.; Yang, W. *J. Chem. Theory Comput.* **2007**, *3*, 390.
(54) Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W. *J. Chem. Phys.* **2008**, *128*, 034105.
(55) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
(56) Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265.
(57) Wilson, E. B., Jr.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
(58) Peng, C.; Ayala, P. Y.; Schlegel, H. B. *J. Comput. Chem.* **1996**, *17*, 49.
(59) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
(60) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(61) Mo, Y.; Gao, J. *J. Phys. Chem. A* **2000**, *104*, 3012.
(62) Chuang, Y.-Y.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quantum Chem.* **1998**, *70*, 887.
(63) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(64) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(65) Thiel, W. *MNDO97 v5.0*; University of Zurich: Zurich, Switzerland, 1998.
(66) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(67) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khaliullin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.,

III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
(68) Carroll, D. L. *Genetic Algorithm Driver 1.70*; University of Illinois: Champaign, IL, 1998.
(69) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.
(70) Hoover, W. *Phys. Rev. A* **1985**, *31*, 1695.
(71) Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136.
(72) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327.
(73) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.
(74) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(75) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(76) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(77) Miller, W. H.; Handy, N. C.; Adams, J. E. *J. Chem. Phys.* **1980**, *72*, 99.
(78) We noticed that although HF/3-21G/MM gives greater absolute solvation energies for related species (Table 2) compared to the PM3-SRP/MM results (this is perhaps related to the underpolarization tendency of the SE methods as we discussed in the Introduction), it is the difference of the solvation contributions between the reactant and the transition state that determines how the solvation effect modifies the gas-phase energy barrier.
(79) (a) Plotnikov, N. V.; Kamerlin, S. C. L.; Warshel, A. *J. Phys. Chem. B* **2011**, *115*, 7950. (b) Plotnikov, N. V.; Warshel, A. *J. Phys. Chem. B* **2012**, *116*, 10342.
(80) Zhou, Y.; Ojeda-May, P.; Pu, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15811.
(81) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III *J. Phys. Chem. A* **2001**, *105*, 9396.
(82) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807.
(83) Pu, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1192.