

Supporting Information for publication

“Exploring, Refining and Validating the
Paradynamics QM/MM Sampling”

*Nikolay Plotnikov and Arieh Warshel**

Department of Chemistry (SGM418), University of Southern California, 3620 McClintock Ave.,
Los Angeles CA-90089, U. S. A.

*warshel@usc.edu

I. Implementation of solute polarization by MM atoms in MOPAC2009 interfaced with molaris9.11

In the present work all the QM/MM simulations were performed using a development version of MOLARIS simulation package (MM program) and a development version of MOPAC2009¹ (QM program) which we modified to enable the QM/MM coupling described below.

In this QM/MM approach, the electrostatic coupling is implemented according to earlier works^{2,3} (See Eq. 34 of the main text). The implementation involved:

Evaluation by the MM program of the electrostatic potential (ESP), vqc , at centers of the QM atoms:

$$vqc(i) = 332 \sum_{\text{MM atoms } j} \frac{q_j}{r_{ij}} \quad (\text{S1})$$

Where i is the QM atom number, j is the MM atom number, r_{ij} is the distance between the atoms i and j , $[vqc] = \text{kcal}/(\text{mol} \cdot e)$ where e stands for the elementary charge.

The QM program reads the supplied ESP from the file mol.in created by the MM program:

```
first_line
6 0 # of qmmm atoms, # of link atoms in Region I
CL -1.591010336 -3.497323620 -4.177329152 119.381953977
C 0.623273531 -3.927769978 -4.243650888 88.802327810
H 0.627631085 -3.831528682 -5.334074435 77.449540155
H 0.737788528 -3.010768158 -3.634868517 83.899739734
H 0.444587282 -4.863821218 -3.677635261 90.477795343
CL 2.837655032 -4.254371189 -4.197078072 120.024810232
```

The reading is activated by (DEBUG + MOL_QMMM(new)) keywords, e.g.:

```
PM6 ISCF CHARGE=-1 SINGLET GRAD LET XYZ DEBUG MOL_QMMM
snapshot of MD step 0
```

```

CL -1.5910103360 1 -3.4973236200 1 -4.1773291520 1
C 0.6232735310 1 -3.9277699780 1 -4.2436508880 1
H 0.6276310850 1 -3.8315286820 1 -5.3340744350 1
H 0.7377885280 1 -3.0107681580 1 -3.6348685170 1
H 0.4445872820 1 -4.8638212180 1 -3.6776352610 1
CL 2.8376550320 1 -4.2543711890 1 -4.1970780720 1

```

These new keywords make the QM program open the file mol.in. The first line is skipped; on the second line the first two numbers (QM atoms + link atoms) are read to determine the total number of atoms in the QM input. $vqc(i)$ on the link atoms can be approximated by the ESP on the host QM atom. We add to the one-electron matrix elements the energy of interaction between the electron and MM atoms, $-vqc(i)$:

E.g. to convert kcal/mol into eV:

$$h_{\text{QM/MM}}(i) = h_{\text{QM}}(i) - \frac{vqc(i)}{23.0606} \quad (\text{S2})$$

We have made these few changes to the QM program. The MM-program now reads energy (heat of formation), gradients and the solute charges (note that now the solute wavefunction is polarized and the calculated heat of formation and the solute charges correspond to the polarized solute). Thus, the QM/MM coupling electrostatic energy is included in the calculated heat of formation. However, the electrostatic QM/MM coupling in the energy gradients (both for QM and for MM atoms) is still needed to be calculated by the MM program using the read solute charges.

Note that the MOPAC2009 program is developed and distributed by Stewart Computational Chemistry, and authors of this article have worked with the MOPAC2009 code under the non-

distribution code-donor agreement. The MOLARIS⁴ program is developed in Arie Warshel's research group (laetro.usc.edu) and might be available on request.

II. Computational details

Simulation 1. FEP from PM3/MM to PM6/MM for the S_N2 reaction in water between methyl chloride and chloride

The FEP between the 2 QM/MM potentials, E_{REF} (PM3/MM) and E_{TARG} (PM6/MM) at the TS (transition state), was done between the potentials given by Eq. 16 and 17 of the main text. Note that the solute charges in these simulations are derived according to the scheme of Mulliken. The system was solvated by an 18 Å sphere of water⁵ and pre-equilibrated on the potential E_{REF} . A harmonic constraint with $K=100$ kcal/(mol·Å²) was used to contain sampling within the TS ($\xi = d(C-Cl_1) - d(C-Cl_2)$; RC=0 here). We started from 41 independent MD trajectories which were propagated for 100,000 MD steps at 300 K in NVT ensemble (MD step size of 0.5 fs) on one of the mapping potentials given by:

$$E_m = (1 - \lambda_m) E_{PM3/MM} + \lambda_m E_{PM6/MM} + E_{CONS} \quad (S3)$$

where

$$E_{QM/MM} = \langle \Psi_{polar} | H'_{QM/MM} | \Psi_{polar} \rangle + E_{VdW} \quad (S4)$$

and with the weight of the potential E_{TARG} being changed in the equal increments of 0.025 from 0 to 1. Thus, the potentials in the adjacent windows are close which in turn guarantees good overlap between them and good FEP convergence.

The whole procedure was repeated for the simulation at the RS (reactant state) with the only difference that the sampling was contained at RC=-1.75. Note, however, that use of constraints on the RC here is solely for demonstration and testing purposes, since the RS is effectively sampled on the unbiased potential.

Next, we process the MD trajectories to calculate the FEP. Each 10 MD steps we collected the following information: E_{REF} , E_{TAR} , the value of the constraining potential, E_{CONS} , and the two distances, $d(C-Cl_1)$ and $d(C-Cl_2)$ for the QM subsystem. Next, the second half of each MD trajectory (5000 points per simulation window) for the RS was used in the analysis described below. The same procedure was repeated for the simulation at the TS. Two input files obtained in this way were processed by several computational techniques for evaluation of FEP as well as by the generalized FEP/US to get the free energy functions.

Part I: FEP estimates by LRA, FEP and TDI

LRA:

$$\Delta\Delta G_{LRA} = \frac{1}{2} \left(\langle E_{TARG} - E_{REF} \rangle_{E_{REF}} + \langle E_{TARG} - E_{REF} \rangle_{E_{TARG}} \right) \quad (S5)$$

The 1-st and the 2-nd terms (averages of the energy gap between the reference potential (RP) and the target potential (TP), taken correspondingly on the MD trajectories propagated on the RP and on the TP) of Eq. S5.

average FEP:

$$\Delta\Delta G_{FEP} = \frac{1}{2} \left(\sum_{m=1}^{n-1} -kT \ln \left\langle \exp \left(-\frac{E_{m+1} - E_m}{kT} \right) \right\rangle_{E_m} - \sum_{m=2}^n -kT \ln \left\langle \exp \left(-\frac{E_{m-1} - E_m}{kT} \right) \right\rangle_{E_m} \right) \quad (S6)$$

of the forward FEP and the backward FEP which are the 1-st and the 2-nd terms of Eq. S6

n-steps LRA where each step of the full FEP is approximated by LRA:

$$\Delta\Delta G_{LRA}^n = \sum_{m=1}^{n-1} \frac{1}{2} \left(\langle E_{m+1} - E_m \rangle_{E_m} + \langle E_{m+1} - E_m \rangle_{E_{m+1}} \right) \quad (S7)$$

TDI (Thermodynamic Integration) with the trapezoid rule:

$$\Delta\Delta G_{TDI} = \int_0^1 \frac{\partial G}{\partial \lambda} d\lambda = \sum_{m=2}^n \frac{1}{2} \left(\langle E_{TARG} - E_{REF} \rangle_{E_m} + \langle E_{TARG} - E_{REF} \rangle_{E_{m-1}} \right) \Delta\lambda \quad (S8)$$

The aforementioned techniques of Part I obtained from the trajectories generated in Simulation 1 are plotted in Figure 2 of the main text (A) and (B).

Next part of the analysis included:

Part II. FEP/US, constructing the free energy functions

The free energy profiles along $\xi=d(C-Cl_1)-d(C-Cl_2)$ were constructed by:

$$\Delta g_i(\xi) = \Delta\Delta G_m - kT \ln \left\langle \delta(x - \xi) \exp\left(\frac{E_i - E_m}{kT}\right) \right\rangle_{E_m} \quad (S9)$$

Weight-averaged Δg_{REF} profile and the weight-averaged Δg_{TARG} profile over all points were obtained by averaging Eq. S9 over all mapping potentials (simulation frames) with the weighting coefficient:

$$\Delta \bar{g}_i(\xi) = \sum_{frames} \frac{N_j(\xi)}{\sum_{frames} N_j(\xi)} \Delta g_j(\xi) \quad (S10)$$

where $N_j(\xi)$ is the number of times MD visited a particular RC value, ξ , while propagating on the j -th mapping potential, and Δg_j is the free energy function estimate by Eq. S9 on the j -th mapping potential.

Part III. Analysis of MD trajectories and of sampling efficiency

In addition to the above analysis , we created the following histograms:

The distribution of the energy gap between the RP and the TP was obtained by sorting the values from all MD trajectories. This analysis gives the most probable energy gap value, which is a rough estimate of the free energy difference for moving between the two potentials. The distribution of the difference $d(C-Cl_1)-d(C-Cl_2)$ was also calculated. It shows how often the certain values of the RC were visited during the simulation.

The corresponding results of Part III for Simulation 1 are plotted in Figure 3 of the main text. Simulation 1 was also repeated with a larger value of the force constant, $K=1000$, used to contain the sampling within a certain region of the RC.

Simulation 2: PMF's for PM3/MM, PM6/MM potentials

In this series of simulations we used harmonic constraints on the RC to improve the MD sampling. Afterwards we combined the simulation windows and removed the bias potential using the WHAM⁶ program.

We started with the PM3⁷/MM and PM6⁸/MM potentials (Eq. S4) with the QM/MM coupling scheme of Eq. (S2). The range of the Reaction Coordinate (RC) $\xi \in [-2.5, 2.5]$ defined as:

$$\xi = d(C - Cl_1) - d(C - Cl_2) \quad (S11)$$

was divided by the equal increment of $\Delta\xi=0.05$ Å into 101 simulation frames. To contain the sampling within a given simulation window (at the corresponding RC) we used harmonic constraints with the force constant of 100 kcal/(mol·Å²):

$$E_{CONS} = K (\xi - \xi_0)^2 \quad (S12)$$

Starting from the optimized RS gas phase geometry (dipole complex of CH₃Cl and Cl⁻) we changed the distances $d(C-Cl_1)$ and $d(C-Cl_2)$ thus moving the system to an approximate RC value. Next, the obtained geometries were equilibrated in the gas phase using the constraint of Eq. S12 for 4000 MD steps with 0.2 fs step size at 273 K. These equilibrated geometries were solvated with 18 Å spheres of water; the whole system was equilibrated on the actual QM/MM potential with the corresponding constraint for 20000 MD steps (0.2 fs each) at 273 K. The list of non-bonded pair interactions was updated every 5 steps. Next the temperature was increased to 300 K, and the step size was increased to 0.5 fs. The MD trajectories were propagated on the given surface for 50000 steps (with frequency of updating the non-bonded pair interactions equal to 10 steps). After 10000 MD steps we started collecting data for PMF calculations each 10 steps, which were performed using the WHAM⁶ program. The PMF's constructed for the PM3/MM potentials (with the charge derivation schemes by Mulliken and by the ESP-fitting) and for PM6/MM potential (with the Mulliken charge derivation) are given in Figure 5 of the main text.

Simulation 3: FEP from EVB to PM3/MM

In this simulation we used the EVB potential as a RP and PM3/MM (Mulliken charges) as a TP. The parameters for the EVB potential are the refined parameters which can be found elsewhere⁹. In this work we just used a slightly different EVB off-diagonal element:

$$H_{12} = 38 \exp(-1.05 \cdot \xi^2) \quad (\text{S13})$$

also, the bond-angle coupling was on in all calculations and the C...Cl- interactions were additionally described by the harmonic constraint in both EVB states (this energy was included in the EVB diabatic states in kcal/mol):

$$E(r(C...Cl^-)) = 5(r-3)^2 \quad (\text{S14})$$

The EVB RP is given by:

$$E_{EVB} = c_1^2 E_1 + c_2^2 E_2 - 2c_1 c_2 H_{12} \quad (\text{S15})$$

while the PM3/MM TP is given by Eq. S4. We used a harmonic constraint at the RC=-1.5 (which corresponds to the RS) to check the effect of the constrained on the FEP at a particular RC with the LRA treatment. Other simulation details were similar to the ones described in Simulation 1. At the EVB TS, we found it more efficient to constraint the eigenvectors component to be equal (which means mixing 2 EVB diabatic states equally, note that for the EVB TS $E_1=E_2$ and thus $c_1=c_2$), i.e.:

$$E_{EVB\neq} = \frac{1}{2} E_1 + \frac{1}{2} E_2 - H_{12} \quad (\text{S16})$$

That is, we carried out the perturbation from the potential given by Eq. S16 to the PM3/MM potential, with sampling still being contained within the TS by the harmonic constraint at RC=0 as in Simulation 1.

The data points were collected every 10 MD steps after discarding the first 50000 steps for equilibration. While for the trajectory propagated at the RS we used exactly the same formalism as described in Simulation 1 (created the input file 41x5000 points, and performed Part I-Part III analysis of the trajectories), for the TS we had to modify the procedure of constructing the free energy functions. Namely, we calculated the actual EVB adiabatic energy (since, while propagating the MD trajectory on the potential of Eq. S16, E_1 can be different from E_2 even when the eigenvector components are forced to be equal):

$$E_{EVB} = \frac{1}{2} \left[E_1 + E_2 - \sqrt{((E_1 - E_2)^2 + 4H_{12}^2)} \right] \quad (S17)$$

Subsequently, we constructed the free energy functions by Eq. S9 and S10.

Note that the results displayed in Figure 6 of the main text were obtained by Eq. S5-S8 for both FEP (i.e. E_{TARG} is given by Eq. S16 at the TS), as were the results presented in Figure 7. However, the results displayed in Figure 8 were obtained by completely removing the actual bias potential which is just the harmonic constraint at the RS, but additionally includes the difference between the Eq. S16 and Eq. S17 for the TS. In Figure 9 the free energy function at the TS for the potential of Eq. S16 is given for comparison (i.e. only the bias due to the harmonic constraint was removed).

Finally, to construct a PMF for the EVB potential (the PMF for the PM3/MM potential was constructed in Simulation 2) we performed an EVB FEP/US calculation. That is we used the EVB mapping potentials to sample the EVB potential along the reaction coordinate defined as the energy gap between the EVB diabatic states, E_1 and E_2 , also frequently called the product state and the reactant state:

$$E_m = (1 - \lambda) E_1 + \lambda E_2 \quad (\text{S18})$$

In total we generated 41 simulation frames, and ran them sequentially, for 50000 MD steps each (0.5 fs step size, 300K, non-bonded pair list updated every 5 steps, 18 Å simulation sphere of water). Before the first EVB frame we let the system equilibrate for 10000 steps.

For each frame after the first 10000 steps we collected data (each 10 steps) to construct the free energy function using the FEP/US approach (4000 per frame). Then the data points were processed using Eq. S9 what gave the results reported in Figure 11 of the main text.

Fitting of Γ -functions to the Potential Energy Scans

First, we performed the potential energy scan along the RC using the QM program for the methyl chloride and chloride system. The scans were carried out at PM3, PM6 level of theory in the gas phase and in the COSMO implicit solvent model.

Second, we used the approach described in Simulation 2, which involved propagating 101 MD trajectories on the EVB potential in the gas phase with additional harmonic constraints with the force constants, K , of 70 kcal/(mol·Å²) imposed on the equally spaced values of the RC in the range $\xi \in [-2.5, 2.5]$. Then we constructed the PMF's for the EVB potentials (using the WHAM

approach) with the constraints of the Eq. S14 acting in the full range of the RC and in the range where the constraints were imposed only within the RC region of $\xi \in [-1.0, 1.0]$.

Finally, the obtained PES's and the PMF were fitted by the sums of Gaussians by searching the minimum of the least square functions given in Eq. 25 of the main text. We found that it was convenient to perform the fitting with a mathematical software package Maple9.5, which gives the analytical derivatives of Gaussians (which were subsequently used in the optimal steepest descent minimization approach to determine the parameters of the Gaussian functions), another advantage being a convenient graphical user interface. See the script example given below.

The data points of the PES's and of the PMF used for fitting, as well as the derived Γ -functions (the sum of the Gaussian functions approximating the corresponding energy functions obtained in the fitting procedure) are given in Figure 11 of the main text. Additionally, the parameters of Gaussians are reported in Table 1 of the main text.

Simulation 4: Gas-phase EVB PMF for the original EVB potential and for the refined EVB potential

In this simulation, we additionally calculated the PMF for the refined EVB potential in the gas phase (the PMF for the original EVB potential and the computational details are described in Fitting of Γ -functions to the Potential Energy Scans):

$$E'_{EVB} = E_{EVB} - \Gamma_{EVB} + \Gamma_{PM6} \quad (\text{S19})$$

The comparison of the free energy functions for the original EVB potential (S15) and for the refined EVB potential (S19) is given in Figure 12.

Simulation 5: AC solvation free energies

The PM3 optimized gas-phase geometries were first evaluated for both the RS and for the TS. Next we calculated the work of charging the solute by Eq. S6 (to get the solvation free energies for the TS and for the RS reported in Table 2 of the main text) from $\mathbf{Q}_0=\mathbf{0}$ to $\mathbf{Q}(\text{RS})$ and $\mathbf{Q}(\text{TS})$ by performing the FEP with 26 mapping potentials of Eq. S18, where

$$E_1 = E_{vdW} \quad (\text{S20})$$

$$E_2 = E_{vdW} + 332 \sum \frac{Q_i q_j}{r_{ij}} \quad (\text{S21})$$

That is, we propagated 26 MD trajectories on the RS (and for the TS) with the frozen solute coordinates. After the initial equilibration of the solute with zero charges in an 18 Å sphere of water for 20000 MD steps (1 fs, 300 K), each trajectory was sequentially propagated for 20000 MD steps (1 fs, 300 K).

| | \mathbf{Q}_0 | $\mathbf{Q}(\text{RS})$ | $\mathbf{Q}(\text{TS})$ |
|--------------|----------------|-------------------------|-------------------------|
| <i>Cl(L)</i> | 0.0 | -0.1706 | -0.7441 |
| <i>C</i> | 0.0 | -0.4322 | 0.1407 |
| <i>H</i> | 0.0 | 0.1974 | 0.1160 |
| <i>H</i> | 0.0 | 0.1969 | 0.1158 |
| <i>H</i> | 0.0 | 0.1972 | 0.1157 |
| <i>Cl(A)</i> | 0.0 | -0.9888 | -0.7441 |

Also for the same geometries we calculated the solvation free energy by the COSMO approach as implemented in the QM program.

The obtained results are reported in Table 2 of the main text.

Simulation 6: The PMF for COSMO and its improvement by the with Γ -COSMO

We started by propagating 101 independent MD mapping trajectories for the MeCl +Cl⁻ system in the gas phase. The mapping potentials defined with the equally spaced coupling parameter $\lambda \in [0.0, 1.0]$

$$E_m = (1 - \lambda_m) E_{RS} + \lambda_m E_{PS} \quad (\text{S22})$$

where

$$E_{RS} = E_{QM} - \Gamma_{QM} + K (\xi - \xi_{RS})^2 \quad (\text{S23})$$

$$E_{PS} = E_{QM} - \Gamma_{QM} + K (\xi - \xi_{PS})^2$$

$QM=PM3+COSMO$; $K=5 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$; $\xi_{RS} = -\xi_{PS} = -2.5 \text{\AA}$ RC is defined by Eq. S11

Each trajectory was initially equilibrated during 20000 MD steps (0.3 fs step size, 273 K) and then was propagated for 100000 steps (1 fs step size, 300K), during the final 75000 steps the data points were collected, each 30 steps, for the subsequent free energy calculations. Additionally, we used a harmonic constraint on Cl₁-C-Cl₂ angle (at π rad with a force constant, $K=15 \text{ kcal}/(\text{mol} \cdot \text{rad}^2)$). All the computer experiments were performed in the gas phase, while the solvent effect was taken into account by the implicit COSMO model as implemented in MOPAC2009.

In the free energy calculations we used Eq. S9 where E_m is given by Eq. S22 and i =PM3+COSMO (PM3 potential with the COSMO solvent, ϵ =78.4).

To get the PM3/MM free energy function for the RC we used $E_{QM}=E_{PM3/MM}$ given by Eq. S4. The simulation protocol was following. First, by changing $d(C-Cl_1)$ and $d(C-Cl_2)$ we moved the system near the region of interest of the RC. Subsequently, we relaxed the system on each potential specified by Eq. S22 in the gas phase with PM3+COSMO for 25000 MD steps (0.3 fs stepsize, 273 K, $K=10$ in Eq. S23). At the next step we solvated the relaxed geometry by an 18 Å sphere of water molecules and relaxed on the PM3/MM potential (with the ESP-derived solute charges) for 25000 MD steps (at 273 K, with a step size 0.3 fs and the frequency of updating the non-bonded pair list of 5, $K=10$). Finally, we increased the temperature to 300K and ran a 100000 steps MD trajectory with a step size of 1 fs (10 nonbond update frequency, $K=10$) on the same potential. The data points for the free energy calculations were collected each 30 steps for the last 2/3 of the trajectory. In the free energy calculations we used Eq. S9 where E_m is given by Eq. S22 and i =PM3/MM with the ESP-derived solute charges.

The calculated free energy profiles are given in Figure 13 of the main text. Additionally, we explored the sampling efficiency for the PM3/MM and PM3+COSMO run described here by the approach described in Simulation 1 (Part III). That is, we constructed the distribution of the sampled RC values for the point used in the free energy analysis. See Figure 14 of the main text.

Simulation 7: The free energy functions obtained using the EVB-type solvent driving potentials. Refining a MO-based QM/MM reference potential

The approach of Eq. S22 and S23 was extended by including an additional term in the mapping potential:

$$E_{ss}(\mathbf{Q}_k) = \sum_{i=1}^{QM} \sum_{j=1}^{MM} \left(\frac{Q_i^{(k)} q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (S24)$$

This was done by using :

$$E_{map} = E_{PM3, gas} - \Gamma_{PM3, gas} + (1 - \lambda)(E_{CONS}(\xi_{RS}) + E_{ss}(\mathbf{Q}_{RS})) + \lambda(E_{CONS}(\xi_{PS}) + E_{ss}(\mathbf{Q}_{PS})) \quad (S25)$$

where \mathbf{Q}_{RS} and \mathbf{Q}_{PS} are the vectors of the QM charges at the RS and at the PS (ESP-charges for the PM3 gas-phase optimized RS):

| | Cl _L | C | H | H | H | Cl _A |
|------------------------|-----------------|--------|-------|-------|-------|-----------------|
| $\mathbf{Q}_{PM3}(RS)$ | -0.180 | -0.382 | 0.183 | 0.183 | 0.183 | -0.987 |
| $\mathbf{Q}_{PM3}(PS)$ | -0.987 | -0.382 | 0.183 | 0.183 | 0.183 | -0.180 |

First, by changing $d(C-Cl_1)$ and $d(C-Cl_2)$ we moved the system to an approximate value of the RC. Then we relaxed the system on each potential derived according to Eq. S22 in the gas phase with PM3 gas phase potential for 2000 MD steps (0.2 fs stepsize, 273 K, $K=10$ in Eq. S23). Second, we solvated the relaxed geometry by an 18 Å sphere of water molecules and relaxed on the potential given by Eq. S25 for 10000 MD steps (at 273 K, with a step size 0.2 fs and the frequency of updating the non-bonded pair list of 5, $K=10$). Finally, we increased the temperature to 300K and ran a 50000 steps MD trajectory with a step size of 1 fs (10 nonbond update frequency, $K=10$) on the same potential. The data points for the free energy calculations were collected each 10 steps for the final 3/5 of the trajectory.

In the free energy calculations we used Eq. S9 where E_m is given by Eq. S25 and the PM3/MM potential is approximated by the classically coupled with MM, according to Eq. S24, gas phase PM3 (solute charges are derived by Mulliken)

$$E_i = E_{\text{PM3,gas}} + E_{\text{ss}}(\mathbf{Q}_{\text{PM3,gas}}) \quad (\text{S26})$$

The results of the free energy calculation are presented in Figure 15 (A) where the comparison with the PM3/MM potential obtained in Simulation 2 is made.

In the next step we repeated the free energy calculations (with the same data generated on potential of Eq. S25) using for the E_i of Eq. S9 the following expression :

$$E_i = E_{\text{PM3,gas}} + E_{\text{ss}}(\mathbf{Q}_{\text{PM3,gas}}) - \Gamma_{\text{PM3,gas}} + \Gamma_{\text{PM6,gas}} \quad (\text{S27})$$

The results of this free energy calculation are presented in Figure 15 (B) where the comparison with the PM6/MM potential obtained in Simulation 2 is also made.

Simulation 8: Refinement of the EVB reference potential

Finally, we refined the EVB potential (used in Simulation 3) in the following way:

$$E'_{\text{EVB}} = E_{\text{EVB,gas}} + c_1^2 E_{\text{ss}}(\mathbf{Q}_{\text{PM3+COSMO}}(RS)) + c_2^2 E_{\text{ss}}(\mathbf{Q}_{\text{PM3+COSMO}}(PS)) - \Gamma_{\text{EVB,gas}} + \Gamma_{\text{PM3,gas}} \quad (\text{S29})$$

$$E''_{\text{EVB}} = E_{\text{EVB,gas}} + c_1^2 E_{\text{ss}}(\mathbf{Q}_{\text{PM6+COSMO}}(RS)) + c_2^2 E_{\text{ss}}(\mathbf{Q}_{\text{PM6+COSMO}}(PS)) - \Gamma_{\text{EVB,gas}} + \Gamma_{\text{PM6,gas}} \quad (\text{S30})$$

where $E_{\text{EVB,gas}}$ describes the intra-interactions within the solute of the EVB potential.

Mulliken charges for the PM3+COSMO optimized RS:

| | Cl _L | C | H | H | H | Cl _A |
|----------------------------|-----------------|--------|-------|-------|-------|-----------------|
| $Q_{\text{PM3+COSMO}}(RS)$ | -0.137 | -0.067 | 0.066 | 0.066 | 0.066 | -0.994 |
| $Q_{\text{PM3+COSMO}}(PS)$ | -0.994 | -0.067 | 0.066 | 0.066 | 0.066 | -0.137 |

Mulliken charges for the PM6+COSMO optimized RS:

| | Cl _L | C | H | H | H | Cl _A |
|----------------------------|-----------------|--------|-------|-------|-------|-----------------|
| $Q_{\text{PM6+COSMO}}(RS)$ | -0.226 | -0.319 | 0.181 | 0.181 | 0.181 | -0.998 |
| $Q_{\text{PM6+COSMO}}(PS)$ | -0.998 | -0.319 | 0.181 | 0.181 | 0.181 | -0.226 |

The EVB FEP/US calculations were performed using the mapping potentials:

$$E_m = (1 - \lambda) E_1 + \lambda E_2 - \Gamma_{\text{EVB,gas}} + \Gamma_{\text{PM3,gas}} \quad (\text{S31})$$

and

$$E_m = (1 - \lambda) E_1 + \lambda E_2 - \Gamma_{\text{EVB,gas}} + \Gamma_{\text{PM6,gas}} \quad (\text{S32})$$

Each simulation involved 41 frames, run sequentially, for 50000 MD steps each (0.5 fs step size, 300K, non-bonded pair list updated every 5 steps, 18 Å simulation sphere of water). Before the first EVB frame we equilibrated the system for 10000 steps.

For each frame we collected data (each 10 steps ,after the first 10000 steps) to construct the free energy function using FEP/US approach (4000 points per frame). The data points were processed using Eq. S9 for E_i and E_m given by Eq. S29 and S31, and for E_i and E_m given by Eq. S30 and S32. The obtained free energy functions are reported in Figure 16 of the main text, where the comparison with the original EVB potential is also given.

Additional notes

Typically there are about 800 solvent molecules within the distance of 18.0 Å from the solute

Parameters for water:

| | |
|---------------------------|-------------------------------------|
| water | !name of the first solvent |
| 0.9897 | !density (gram/cm**3) |
| 0.99 0.99 | !bond length for bond 1, bond 2 |
| 375.0 375.0 | !force constant for bond 1, bond 2 |
| 1.911 | !angle between bond 1 and bond 2 |
| 60.0 | !force constant for angle |
| 00861.440 026.040 016.000 | !VDW a,b and mass for OW |
| 00000.420 000.575 001.000 | !VDW a,b and mass for W1 |
| 00000.420 000.575 001.000 | !VDW a,b and mass for W2 |
| -0.82 0.41 0.41 | !charges for the three united atoms |

Maple9.5 script to fit the PES with 3 Gaussian functions

#input file format: RC Energy

```
f := fopen( "PM3_PES.csv", READ );
t1:=readdata(f,[float,float]);
fclose(f);

#INPUT PARAMETERS

npoints:=130;           #number of point in PES

A0:=-3.8: print("A=",A0);  #guess value for the height of Gaussians

B0:=5.8: print("B=",B0);

C0:=0:

R1:=-1.0:               #values for the centers of Gaussians

R2:=0:

R3:=1.0:

niter:=20:

#THE END
```

```

temp:=C0:

C0:=temp+A0:

fvalue:=array(1..npoints+1):

counter:=array(1..npoints+1):

Eref:=array(1..npoints):

Efit:=array(1..npoints):

xi:=array(1..npoints):

Eref:=array(1..npoints):

print("PASS1");

for j from 1 to npoints by 1 do

    xi[j]:=t1[j,1];

    Eref[j]:=t1[j,2]-t1[1,2];

end do:

a0:=0.5:

b0:=1.0:

g0:=0.5:

print("PASS2");

Ef:=proc( x, alpha, beta, gamma , A1, B1, C1)

    description "evaluate _PDMTD";

    A1*exp(-alpha*(x-R1)^2)+B1*exp(-beta*(x-R2)^2)+C1*exp(-gamma*(x-R3)^2)

end proc:

print("TROUBLE? npoints=",npoints,"i2=",i2);

F:=sum((Eref[i2]-Ef(xi[i2],a,b,g,A,B,C))^2,i2=1..npoints);

subs(a=a0, b=b0, g=g0, A=A0, B=B0, C=C0, F):

evalf(%);

dFa:=diff(F,a):

dFb:=diff(F,b):

dFg:=diff(F,g):

dFA:=diff(F,A):

dFB:=diff(F,B):

```

```

dFC:=diff(F,C):

print("PASS3");

counter[1]:=1:

fvalue[1]:=subs(a=a0, b=b0, g=g0, F):

for iter from 1 to niter do

    F_old:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, F):

    Gr_a:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFa):

    Gr_b:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFb):

    Gr_g:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFg):

    Gr_A:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFA):

    Gr_B:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFB):

    Gr_C:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, dFC):

    temp_a:=a0: temp_b:=b0: temp_g:=g0: temp_A:=A0: temp_B:=B0: temp_C:=C0:

    ss:=1.0e-7:

#    F_new:=F_old+10:

    for k from 1 to 7 while (k < 2 or F_new < F_old) do

        if (k > 1) then

            F_old:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, F):

            temp_a:=a0: temp_b:=b0: temp_g:=g0: temp_A:=A0: temp_B:=B0: temp_C:=C0:

        end if:

        temp_ss:=ss:

        ss:=temp_ss*10:

        a0:=temp_a - ss* Gr_a:

        b0:=temp_b - ss* Gr_b:

        g0:=temp_g - ss* Gr_g:

        A0:=temp_A - ss* Gr_A:

        B0:=temp_B - ss* Gr_B:

        C0:=temp_C - ss* Gr_C:

        F_new:=subs(A=A0, B=B0, C=C0, a=a0, b=b0, g=g0, F):

```

```

        print("k",k,"F_old",F_old,"Fnew",F_new);

    end do:

    a0:=temp_a: b0:=temp_b: g0:=temp_g: A0:=temp_A: B0:=temp_B: C0:=temp_C:

    print("iter",iter,"ss",ss,"F",F_old,"alpha",a0,"beta",b0,"gamma",g0,"A",A0,"B",B0,"C",C0);

    counter[iter+1]:=iter:

    fvalue[iter+1]:=F_old:

end do:

print("PASS4");

xylist := zip ( (x,y) -> [x,y], counter, fvalue):

plot(xylist,style=point,symbol=cross,color=black,title="STEEPEST DESCENT MINIMIZATION CURVE");

print("PASS5");

for i2 from 1 to npoints do

    Efit[i2]:=Ef(xi[i2],a0,b0,g0,A0,B0, C0):

end do:

xylist2 := zip ( (x,y) -> [x,y], xi, Efit);

xylist3 := zip ( (x,y) -> [x,y], xi, Eref);

G0:=plot(xylist2,style=point,symbol=cross,color=green,title="PDMTD");

G1:=plot(xylist3,style=point,symbol=cross,color=red,style=point):

with(plots):

display({G0,G1});

```

References

- (1) Stewart, J. J. P.; MOPAC2009 ed.; Stewart Computational Chemistry: Colorado Springs, CO, 2008.
- (2) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.
- (3) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199-213.
- (4) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161-185.
- (5) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124-129.
- (6) Grossfield, A.; WHAM 2.0.6 ed.; University of Rochester: Rochester, NY, 2012.
- (7) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209-220.
- (8) Stewart, J. *J. Mol. Model.* **2007**, *13*, 1173-1213.

(9) Plotnikov, N. V.; Kamerlin, S. C. L.; Warshel, A. *J. Phys. Chem. B* **2011**, *115*, 7950.