Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

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

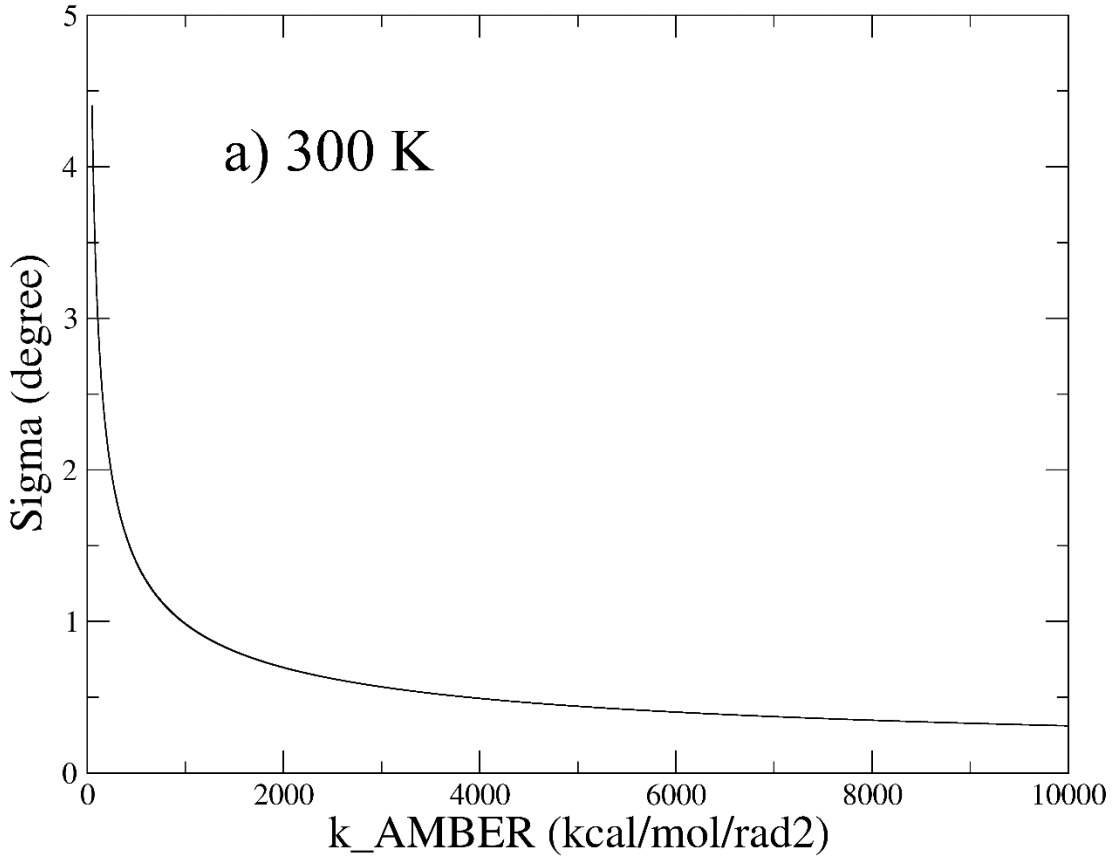
Zhaoxi Sun[1,2*]

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

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

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

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



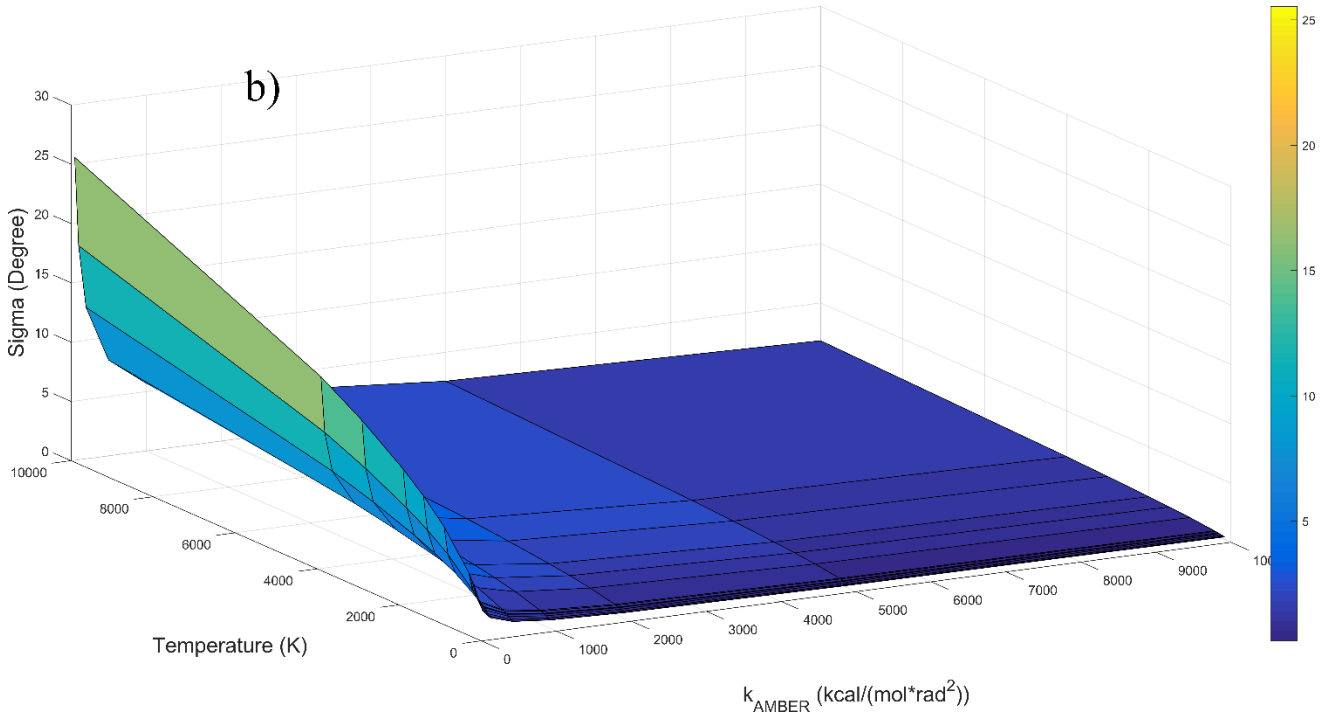
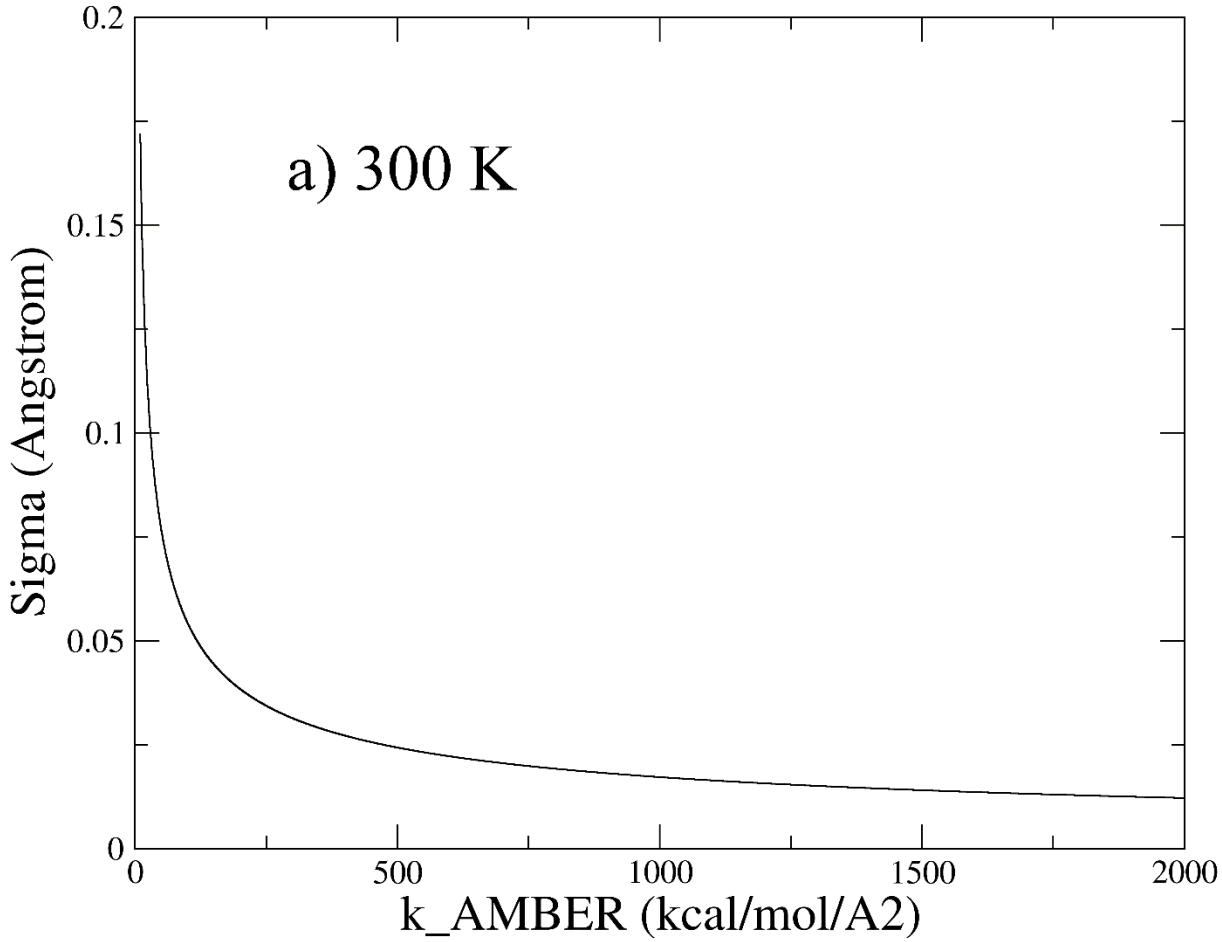


Fig. S2. a) σ as a function of force constant in AMBER at 300 K in the distance case. b) σ varies in the spaces of force constant in AMBER and simulation temperature in the distance case.



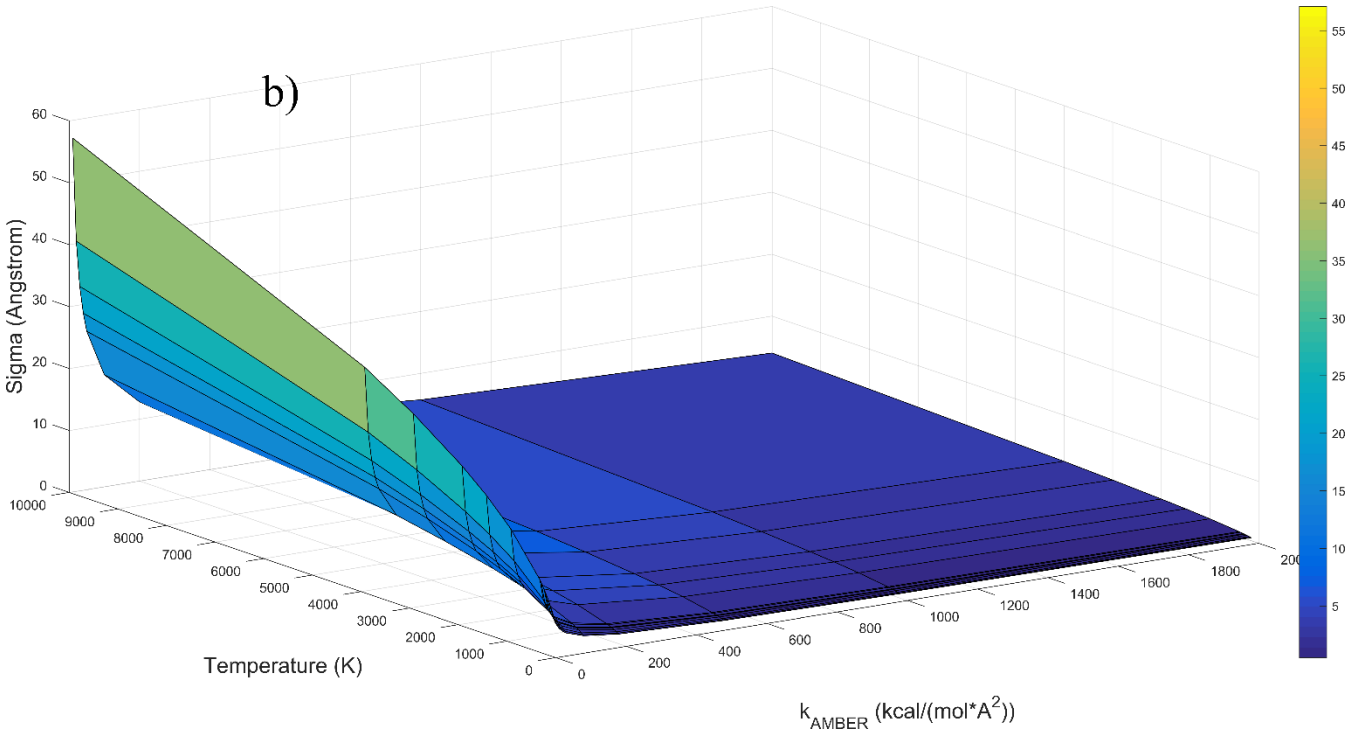


Fig. S3. Under different ab initio QM Hamiltonians, the convergence behavior of free energy profiles the pulling speeds at a) HF, b) MP2, c) ωB97X-D. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. The time in the legend represents the pulling time for each 2° segment. From the pulling-speed dependence, we know that 0.5 ps per segment is slow enough for convergence and thus the statistics under this pulling speed is used for discussion in the following parts of the paper.

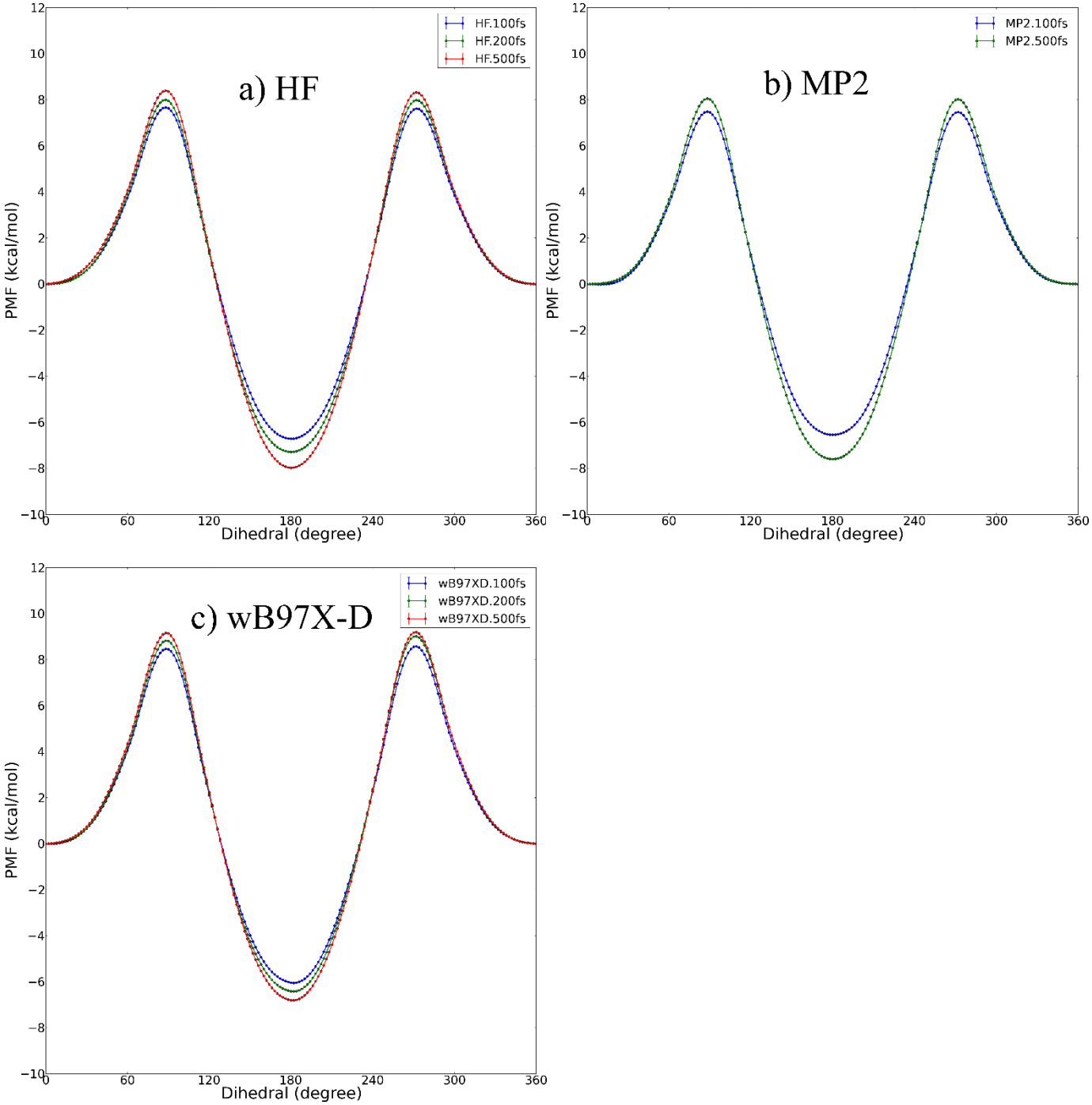
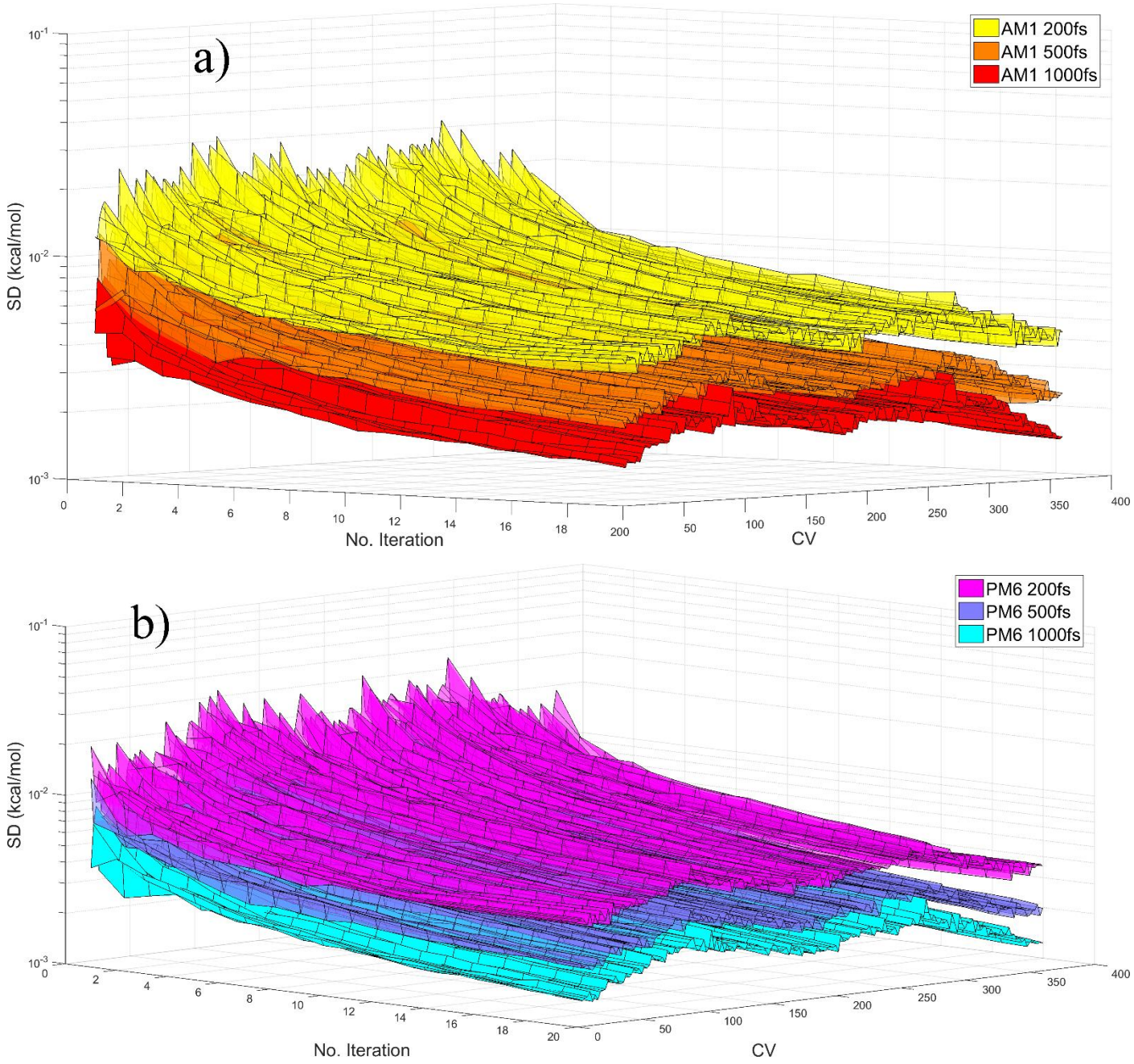
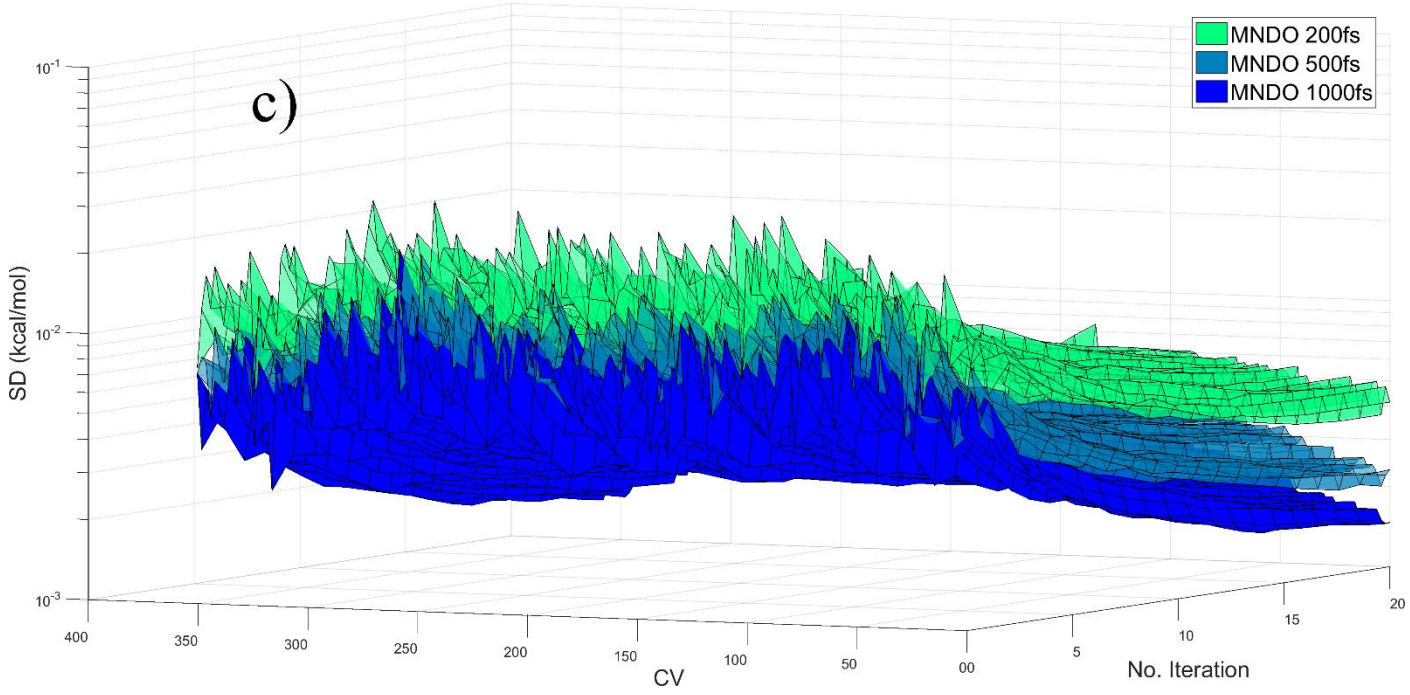


Fig. S4. Sample-size dependence of SD profiles for direct free energy simulation at different levels of theory. a) AM1, b) PM6, c) MNDO, and d) B3LYP.




c)
MNDO 200fs
MNDO 500fs
MNDO 1000fs
SD (kcal/mol)
CV
No. Iteration

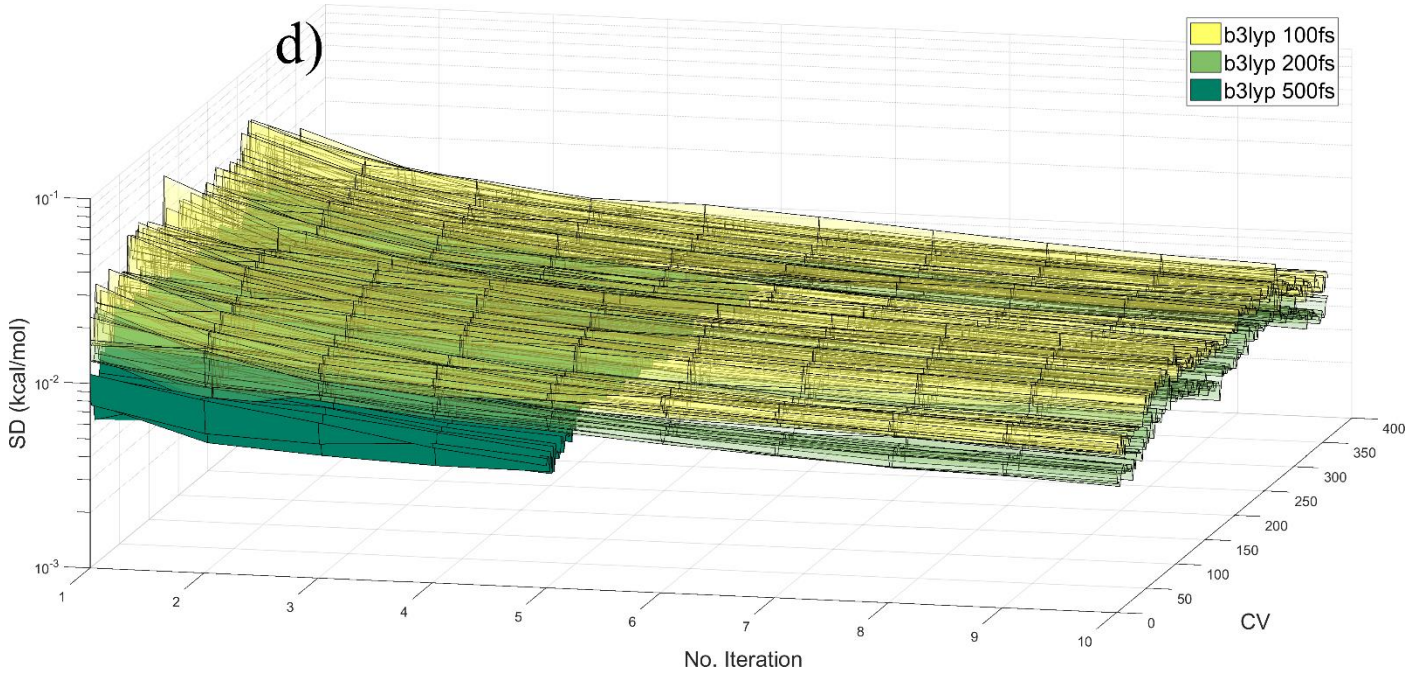

d)
b3lyp 100fs
b3lyp 200fs
b3lyp 500fs
SD (kcal/mol)
No. Iteration
CV

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

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

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

Direct Free Energy Simulations

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

SQM<->QM corrections

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

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

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

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

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

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

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