universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Towards affordable QM/MM free energy simulations“

verfasst von / submitted by

Andreas Schöller BSc MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2023/ Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 796 605 419
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Chemie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Stefan Boresch
Mitbetreut von / Co-Supervisor:	

Abstract(English)

Most free energy simulations using a multiscale quantum mechanics (QM)/molecular mechanics (MM) Hamiltonian use an indirect thermodynamic cycle to avoid performing so-called "alchemical" transformations at the high (QM/MM) level of theory. Nonequilibrium work (NEW)-based free energy calculations have been shown to be a reliable method for computing the correction free energy between the MM and QM/MM description of the intra- and intermolecular interactions in these cycles. However, the protocols utilized in previous proof-of-concept studies are too expensive for most real-world applications.

The goal of the thesis is to reduce the cost of the NEW switching simulations, aiming to find an optimal balance between the number of switches and the length of a single switch. The correction steps for the gas phase and aqueous solution were optimized separately. The overall best protocol for the gas phase is almost 40 times more efficient than the ones used previously.

However, directly using this protocol to NEW switches in aqueous solution may lead to incorrect results. The reasons for the convergence problems were analyzed and mitigated by inserting a hybrid-charge intermediate state. In this approach, the region of interest is still described by the regular force field (hence, calculations are fast), but the partial charges are modified to resemble the charge distribution of the QM/MM level of theory. Specifically, for the semi-empirical QM/MM description used in this work, these modified charges are averaged Mulliken charges. When utilizing this intermediate state, the optimized gas phase protocol also works for solvated systems.

In the third main chapter, we apply the optimized protocols to calculate absolute solvation free energies at a semi-empirical QM/MM level of theory. The model systems are compounds for which poor results were obtained using traditional force field-based free energy simulations. While the results are mixed, the applicability of our protocols to large-scale problems is clearly demonstrated. The thesis concludes with a summary of the findings, a critical discussion of the challenges and problems encountered, and an outlook on potential next steps.

Keywords

Nonequilibrium work switching, SQM/MM, optimization, indirect FES

Abstract(Deutsch)

Die meisten Simulationen der Freien Energie, die einen multiskalen quantenmechanischen (QM)/molekularmechanischen (MM) Hamilton-Operator verwenden, benutzen einen indirekten thermodynamischen Zyklus, um sogenannte 'alchemische' Transformationen auf dem hohen (QM/MM)-Niveau der Theorie zu vermeiden. Es hat sich gezeigt, dass Berechnungen, die auf Nichtgleichgewichtsarbeit (NEW) basieren, eine zuverlässige Methode zur Kalkulation der Freien Korrektur-Energie zwischen MM und QM/MM-Beschreibung der intra- und intermolekularen Wechselwirkungen in solchen Zyklen sind. Allerdings sind die in früheren Proof-of-Concept-Studien verwendeten Protokolle für die meisten realen Anwendungen zu teuer. Das Ziel dieser Arbeit ist es, den Aufwand der NEW Switching-Simulationen zu reduzieren, indem ein optimales Gleichgewicht zwischen der Anzahl der Switche und ihrer Länge hergestellt wird. Die Korrekturschritte für die Gasphase und die wässrige Lösung wurden separat optimiert. Das insgesamt beste Protokoll für die Gasphase ist in etwa 40-mal effizienter als die bisher verwendeten. Die direkte Verwendung dieses Protokolls für NEW-Switche in wässriger Lösung kann jedoch zu falschen Ergebnissen führen. Die Gründe für die Konvergenzprobleme wurden analysiert und durch Einführen eines Hybridladungs-Zwischenzustandes gemildert. Bei diesem Ansatz wird der interessante Bereich immer noch durch das reguläre Kraftfeld beschrieben (daher sind die Berechnungen trotzdem sehr schnell). Die Teilladungen werden so modifiziert, dass sie der Ladungsverteilung des QM/MM-Theorieniveaus ähnlicher sind. Im Speziellen sind die Ladungen, die in dieser Arbeit verwendet und für die semi-empirische QM/MM-Beschreibung modifiziert wurden, gemittelte Mulliken-Ladungen. Unter Nutzung dieses Zwischenzustandes funktioniert auch das optimierte Gasphasen-Protokoll bei solvatisierten Systemen. Im dritten Hauptkapitel wenden wir die optimierten Protokolle zur Berechnung von absoluten Freien Solvatationsenergien auf semi-empirischer QM/MM-Theorieebene an. Die Modell Systeme sind Verbindungen, bei denen mit traditionellen Freien Energie Berechnungen, die auf Kraftfeldern basieren, schlechte Ergebnisse erzielt wurden. Obwohl die Ergebnisse gemischt sind, konnten wir die Anwendbarkeit unserer Protokolle für Berechnungen im großen Maßstab deutlich demonstrieren. Die Arbeit endet mit einer Zusammenfassung der Ergebnisse, einer kritischen Auseinandersetzung der aufgetretenen Herausforderungen und Probleme, sowie einem Ausblick auf zukünftig mögliche Forschungsschritte.

Schlagwörter

Nicht-Gleichgewichtsarbeit, SQM/MM, Optimierung, indirekte Freie Energie Simulationen

Acknowledgement

Univ.-Prof. Mag. Dr. Stefan Boresch, words cannot express my gratitude - thank you for your guidance, wisdom, patience, and support. It was a pleasure and an honor for me to work with such a great scientist and mentor. I am extremely grateful for your kind consideration on so many occasions; From the birth of my daughter through all the inconveniences the pandemic brought along; Thank you for showing, that being a good father and a good scientist do not exclude each other.

Prof. Dr. Henry Lee Woodcock III, the collaboration our department shares with you and your lab, is not an ordinary scientific relationship. For more than a decade Prof. Boresch and you accelerate scientific progress in the field - with undeniable success.

Fiona L. Kearns PhD and Phillip S. Hudson PhD, thank you for introducing me to the world of free energy simulations. It was a pleasure to work with such exceptional beings (both personal and scientific). You deserve my highest gratitude, as I owe a lot to you. Your guidance and advice laid the foundation for the success of my studies.

emer. Univ.-Prof. Dr. Othmar Steinhauser, as the founder of our department, you deserve a special acknowledgment; you shaped our department to its current state, a solid and reputable institution that is incomparable and unique. I will miss our conversations as you spread humanism among the machines. Thank you also for your patience and motivation to discuss Quantum Mechanics; your guidance in this field was indispensable.

Univ.-Prof. Dipl.-Chem. Dr. Christian Schröder, thank you for your collaboration and advice during our solvent reorientation dynamics experiments. Your help is greatly appreciated.

Department of Computational Biological Chemistry, thank you all for the kind atmosphere and the strong company that is an integral part of this institution. Our discussions not only fueled my progress but will do for many members to come.

Mag. Martin Mähr, thank you for putting so much faith in my abilities as a scientist. You not only made it possible for me to study in the first place, but you also motivated me to explore my potential and never give up.

Emilia Schallgruber-Schöller and Verena Schallgruber, thank you for your love and support. Without your encouragement and motivation, this dissertation would not have been possible.

My Family, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams.

My Friends, Petala, Florian, and all my friends who always had an open ear and supported me - thank you from the bottom of my heart.

.....

Author's signature

Contents

Thesis

Publications

Appendices

List of Tables

List of Figures

Thesis

1. Introduction

1.1 Molecular dynamics and free energy simulations

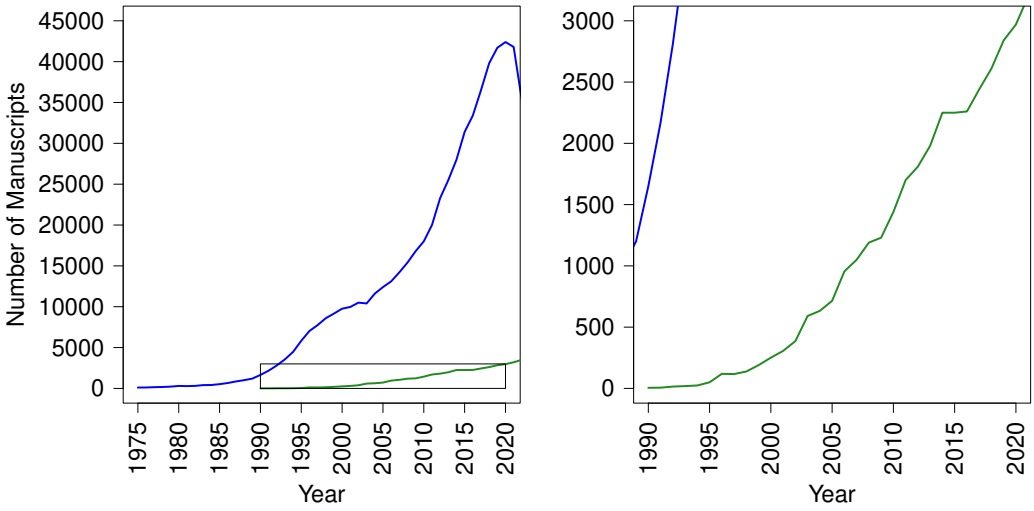


Figure 1.1: Blue: number of published manuscripts on molecular dynamics per year (MD) from 1975 to 2020; green: number of published manuscripts per year on QM/MM in the context of molecular dynamics from 1990 (first appearance) to 2020. Source: Google Scholar

Since the first reported application to a protein in 1977[1] molecular dynamics (MD) simulations have become a versatile tool for studying the properties of biomolecules, complementing experimental techniques [2]. This success is evident from the number of manuscripts published in the field over the last 45 years. Since 1977, the interest of scientists around the world has grown almost continuously, reaching its peak of more than 40,000 published manuscripts in 2020 (see Fig. 1.1)[1]. The central idea of MD simulations is the numerical solution of Newton's equations of motion, which describe the motion of interacting atoms or molecules in a time-dependent manner. To describe the forces between the atoms or molecules, which are needed to solve these equations, the realistic behavior of the system during the simulation is realized by a molecular mechanical (MM) force field (FF). This FF applies rules to the moving particles that are consistent with experimental observations [3]. Since 1977, several molecular mechanical FFs have been introduced and continuously

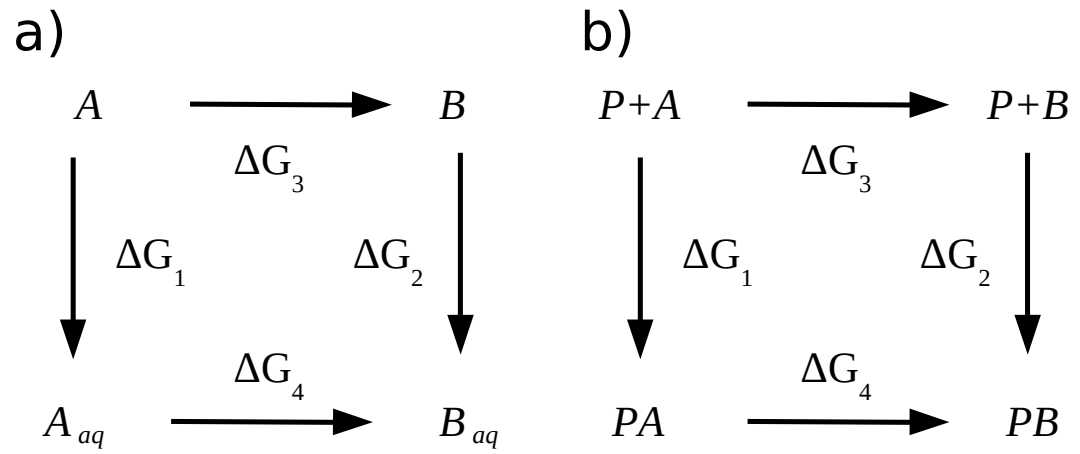
[1] `https://scholar.google.com/`

improved: AMBER[4], CHARMM[5], GROMOS[6], or OPLS[7] to name a few.

Despite the increasing acceptance and use of MD by the community and the continuing development of techniques, intrinsic shortcomings of the methodology have become apparent: a classical FFs operate with fixed point charges and therefore cannot properly describe the electronic structure of a system[8]. In addition, the study of reactions involving the breaking and formation of bonds cannot be performed relying solely on FFs[9]. Other electronic phenomena such as polarizability or charge transfer are not accessible by a classical FF alone, although efforts are being made to include such effects[10, 11, 12]. The comparison of results obtained with the various MM FFs makes clear that their predictive power is strongly limited by their parameterization, leading to strengths and weaknesses of particular FFs to describe particular properties depending on the generation of parameters[13]. These realizations eventually led to the inclusion of quantum mechanics not only in parameter generation but also in simulation.

In 2013, Martin Karplus, Michael Levitt and Arieh Warshel were awarded with the Nobel Prize in Chemistry for the development of the QM/MM method [14]. This methodology allows the combination of quantum mechanics and molecular mechanics in a single simulation. As a kind of parallel evolution, this field also grew steadily, leading to about 3000 publications in the year of 2020 as shown in Figure 1.1. As promising and powerful the application of QM/MM may be, it also has some drawbacks due to its high computational cost and the non-trivial inclusion and treatment of the boundary region between QM and MM[15]. It nevertheless has become a highly used method with increasing popularity over the course of the last decades, which is on its way to become a routinely used method in the simulation of complex biomolecular applications [16].

One potentially extremely useful application of MD is free energy simulations (FES), which allow the calculation of the central quantity of thermodynamics, the free energy[17]. Although the first applications of the methodology were reported around 1990 [18], interest in the methodology in industry was moderate until 2015, when the potential of the method was demonstrated on a comprehensive set of compounds in a case study[19]. Since then, FES has become an increasingly important application of MD and the use of MM FFs. Free energies determine the spontaneity of chemical or biochemical processes and also explain which conformations are thermally accessible. This information is valuable because it allows us to predict whether and how often a biochemical process will occur. It also allows us to determine which reactant is most likely to participate in a reaction when several are present. For this reason, relative free energy differences, which provide a practical way to compare different molecules



(a) Relative solvation free energy (b) Relative binding free energy

Figure 1.2: Typical examples for the use of thermodynamic cycles in free energy simulations; a) Relative solvation free energy; b) Relative binding free energy; in both cases the identity $\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$ holds

regarding this essential property, are of particular interest [20]. Various methods to compute free energy differences between two states have been developed, such as free energy perturbation (FEP)[21], thermodynamic integration (TI)[22], and Bennett acceptance ratio (BAR)[23].

However, the successful calculation of the free energy difference between two molecules is only possible if the two molecules are very similar in size and chemical properties. If they are not, the obtained free energy difference will most likely not be converged. The problem is exacerbated when trying to compare not two, but several molecules with different properties. This problem can be overcome by introducing non-physical intermediate states, which are constructed in such a way that the initial state is gradually morphed into the final one. Such a series of mutations avoids the convergence problems one would encounter when trying to compute the free energy difference between the end states in one step [24, 25].

The power of alchemical FES results not only from being able to calculate the free energy difference between two molecules, but from the combination with suitable thermodynamic cycles [26]. Figures 1.2a and 1.2b illustrate two widely used examples, the calculation of relative solvation free energy differences and the calculation of relative binding free energy differences. Because of its state-function property, the free energy for going around the cycle has to be zero. Therefore, the relative solvation free energy difference $\Delta\Delta G^{solv}_{A\to B} = \Delta G_2 - \Delta G_1$ between a compound A and a compound

B as shown in 1.2a can also be obtained by calculating $\Delta G_{A\to B}$ (ΔG_3), mutating A to B in a simulation without water, and $\Delta G_{A_{aq}\to B_{aq}}$ (ΔG_4), the same transformation but with water present. The same is true for the calculation of relative binding free energies, as shown in 1.2b: The relative binding free energy difference of two ligands, A and B, to a protein P $\Delta\Delta G^{bind}_{A\to B} = \Delta G_2 - \Delta G_1$ can also be obtained by calculating $\Delta G^{bound}_{A\to B}$ (ΔG_4) minus $\Delta G^{unbound}_{A\to B}$ (ΔG_3). If constructed wisely, such cycles allow for a significant reduction in computation.

As mentioned above, the use of an MM FF is sometimes not sufficient to capture all the relevant interactions in the system of interest. For this reason, there is considerable interest in carrying out FES based on QM/MM simulations [27]. An obvious drawback of QM/MM is its high computational cost; therefore, direct calculations of free energy differences at the QM/MM level of theory using the thermodynamic cycles shown in Figure 1.2 is impractical. A less obvious complication is the fact that some tricks to achieve alchemical mutations are not possible with QM/MM [28, 29]. One way around this problem is to construct another thermodynamic circle that allows the desired free energy to be calculated indirectly.

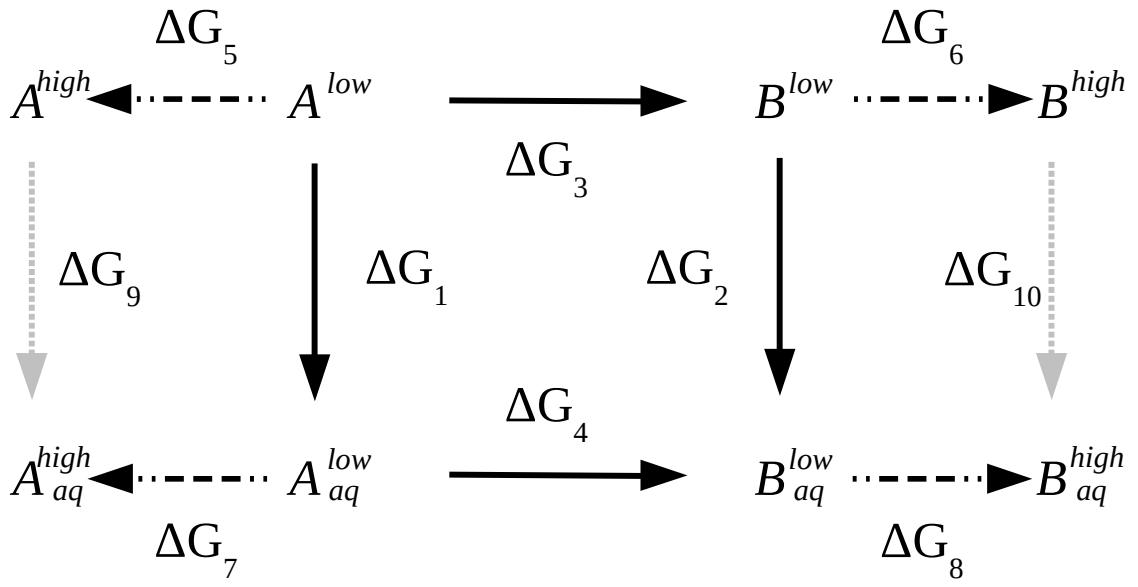


Figure 1.3: Applying end point corrections to a traditional thermodynamic cycle (relative solvation free energy difference) from a *low* to a *high* level of theory.

An example of such an indirect thermodynamic cycle is shown in Figure 1.3. In the example, the goal is to refine the relative solvation free energy difference between A and B calculated at the *low* level of theory, $\Delta\Delta G^{solv\ low}_{A\to B} = \Delta G_2 - \Delta G_1$, to the

high level of theory, $\Delta\Delta G_{A\to B}^{solv\ high} = \Delta G_{10} - \Delta G_9$. The cycle in the middle is the same as in Figure 1.2a; here all simulations are done at the *low* level of theory, i.e., with an MM FF. In contrast to the earlier discussion of thermodynamic cycles, the calculation of the absolute solvation free energies, ΔG_1 ($\Delta G_A^{solv\ low}$) and ΔG_2 ($\Delta G_B^{solv\ low}$), is assumed to be straightforward. In fact, tools like `transformato` make the computation of solvation free energies routine [30].Taking advantage of the fact that the free energy is a state function and considering the leftmost thermodynamic cycle in Figure 1.3, we see that ΔG_9 ($\Delta G_A^{solv\ high}$) can be calculated from ΔG_1 by determining the free energy for the two endpoint corrections, A^{low} to A^{high} (ΔG_5, gas phase) and A_{aq}^{low} to A_{aq}^{high} (ΔG_7, aqueous solution). The solvation free energy of A at the high level of theory, ΔG_9, therefore becomes $\Delta G_9 = \Delta G_1 + (\Delta G_7 - \Delta G_5)$. The same steps must be performed for B, i.e., $\Delta G_{10} = \Delta G_B^{solv\ high} = \Delta G_2 + (\Delta G_8 - \Delta G_6)$. Putting it all together, the relative solvation free energy difference between A and B at the high level of theory, $\Delta\Delta G_{A\to B}^{solv\ high} = \Delta G_{10} - \Delta G_9$ can be calculated as $\Delta G_2 + (\Delta G_8 - \Delta G_6) - [\Delta G_1 + (\Delta G_7 - \Delta G_5)]$.

The derivation outlined above based on Figure 1.3 shows how to refine a relative solvation free energy difference from a low (e.g., MM) to a high (e.g., QM/MM) level of theory. The steps to refine a relative binding free energy (see Figure 1.2b) are analogous. By using the thermodynamic cycle(s) of Figure 1.3, one can, at least in principle, avoid sampling at the high level of theory. The crucial step in such calculations is the accurate calculation of the free energy differences between the two levels of theory (dotted arrows in Figure1.3). The reliable and *efficient* calculation of free energy differences between levels of theory is the focus of this thesis. The detailed analysis of the practical challenges of calculating the low to high corrections is given in Section 1.3.1.

1.2 Calculation of intra- and intermolecular forces

Before proceeding with a more detailed description of how to perform indirect QM/MM FES, we provide an overview of how interactions are calculated at the low and high levels of theory in this work.

1.2.1 Molecular Mechanics (MM)

Today, most biomolecular MD simulations and applications based on them, such as the FES described in the previous chapter, use an additive molecular mechanical FF. The potential energy function includes both intramolecular and intermolecular interactions. The intramolecular contributions to the potential energy consist of terms for bonds, angles, dihedral and improper dihedral angles, and the intramolecular non-bonded interactions. The intermolecular contributions consist of only the non-bonded interactions.

$$\begin{aligned} U = & \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+\cos(n\phi-\delta)] \\ & + \sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2 \\ & + \sum_{nonbonded} \left\{ \epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\} \end{aligned} \tag{1.1}$$

Equation 1.1 represents the potential energy function of the additive CHARMM36 force field[31], which is used in all of our studies. It is the sum of the individual contributions to the total potential energy of the system. Each term represents a property that contributes to the potential energy. The first term in Equation 1.1 describes the bond stretching between atoms, where b is the actual length of a given bond at a given time and b_0 is its equilibrium distance. The square of the difference, which represents the deviation from the equilibrium length, is multiplied by a force constant k_b, which determines the steepness of the potential well. The second term describes the angle between three atoms, including the actual angle θ, its reference angle θ_0 and the angular force constant k_θ. The third term, the dihedral energy term, describes the torsional energy around the central bond between four atoms as a function of the dihedral (or torsional) angle ϕ. Here, k_ϕ is the dihedral force constant, n is the multiplicity, and δ is the phase shift. The fourth term is the so-called improper dihedral energy, which characterizes the out-of-plane angle ω of a fourth atom relative to the plane formed by the first three atoms. As in

the analogous harmonic bond stretching and angle bending terms, k_ω is the force constant and ω_0 is the reference angle. The so-called Urey-Bradley term (fifth term in Equation 1.1) modifies the bending angle between three atoms with respect to the distance u between the end atoms 1 and 3.

The non-bonded interactions are described by the terms in the third row of Equation 1.1. The first two terms of the sum describe the van der Waals interactions between the particles using the Lennard-Jones function. Here r_{ij} is the interparticle distance between atoms i and j, ϵ_{ij} and $R_{min,ij}$ are the Lennard-Jones well-depth and the distance of minimum energy for the atom types to which the two particles belong. The last term describes Coulomb interactions between two particles i and j in terms of their partial charges q. Depending on the FF used, additional terms may be present, often describing a specific aspect of the potential. E.g., the CHARMM protein force field uses the CMAP correction potential, which improves the sampling of the protein backbone angles by a 2D dihedral energy grid correction[32].

The calculation of energies and forces using Equation 1.1 is very efficient, orders of magnitude faster than a quantum chemical description of the interactions. Furthermore, force field terms can be improved incrementally and, at least to some extent, individually. The force field parameters ($k_b, b_0, k_\theta, \ldots, \epsilon_{ij}$, etc.) are obtained by careful parameterization, based either on experimental data, quantum mechanical calculations, or empirical factors that better reproduce the experiment[33]. If one needs to describe molecules that are not part of the standard force fields, then this non-trivial step must be performed by the user.

Despite their advantages, FFs have several limitations. Obviously, they cannot be used to study chemical reactions, as bonds cannot be formed or broken. Even for seemingly simple applications, the neglect of induced electronic polarization can result in incorrect relative free energies of hydration of mono- or bivalent ions [34, 35]. FFs typically omit higher-order terms and important couplings, leading to significant deviations from quantum mechanical representations in geometry (bond lengths, bond angles, and improper dihedral angles), torsional profiles, or vibrational frequencies[36]. In particular, problems are likely for small molecules with multiple chemical moieties, as well as structures including metal ions.

1.2.2 Semi-empirical Quantum Mechanics (SQM) – the SCC-DFTB method

While the title of this thesis might suggest that ab initio QM methods were used as the high level of theory, full quantum chemical calculations (Hartree-Fock, DFT, or better) would have been by far too costly. Instead, we employed a semi-empirical quantum mechanical (SQM) method, self-consistent charge density functional tight binding (SCC-DFTB), which we briefly describe in this section. In fact, many QM/MM studies utilize an SQM rather than a full ab initio QM methods for performance reasons [37]. The reduced computational cost of SCC-DFTB compared to, e.g., full DFT methods made it a reasonable choice to study free energy differences between levels of theory.

SCC-DFTB is a well known and widely used SQM method that delivers reasonable descriptions of systems at the fraction of the cost compared to full density functional tight binding (DFT) methods [38, 39, 40]. For a detailed description of the method, we refer the reader to the original publication and its application to biological systems[41, 42]. After its implementation in CHARMM [43], it was constantly used and improved, resulting in the current third order expansion [44] that was used in all of our studies, combined with the 3ob-3-1 parameters[45, 46, 47, 48]. Recently, strategies were developed that facilitate heterogeneous calculation, where both CPUs and GPUs can be used simultaneously [49, 50]. These developments show the continued and increasing interest in using this SQM method.

$$E^{SCC-DFTB} = \sum_{i,a,b} \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2} \sum_{a,b} \Delta q_a \Delta q_b \gamma_{ab} + \frac{1}{3} \sum_{a,b} \Delta q_a^2 \Delta q_b \Gamma_{ab} + E_{rep} \tag{1.2}$$

The total potential energy when using third order SCC-DFTB (DFTB3) is given by Equation 1.2. This functional form is the result of an iterative development that started with the non-self-consistent density functional tight binding (DFTB) approach[51, 52]. In DFTB, the total energy of the system is approximated as

$$E^{DFTB} = \sum_{i,a,b} \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu i} H^0_{\mu\nu} + E_{rep}. \tag{1.3}$$

The first term in Equation 1.3 is the sum over all occupied valance orbitals i obtained from the Kohn-Sham equations. $\hat{H}_0$ is the Hamiltonian operator resulting from an input density ρ_0 as a superposition of atomic charge densities. The n_i are the occupation numbers, i.e., 1 or 2, and the $c_{\mu i}$ are the expansion coefficients of the

molecular orbitals in the LCAO approximation. The core-core repulsions E_{rep} are the sum of one- or two-body potentials and are pairwise and short-ranged.

Taking E^{DFTB} as a starting point, Elstner and coworkers expanded the approach into a self-consistent variant:

$$\sum_b \sum_{\nu \in b} c_{\nu i} \left(H_{\mu\nu} - \epsilon_i S_{\mu\nu} \right) = 0 \;\; \forall a, \mu \in a, i \tag{1.4}$$

Here, $S_{\mu\nu}$ are the overlap integrals of the atomic basis functions. The coefficients $c_{\nu i}$ are optimized self-consistently until Equation 1.4 is fulfilled. The approach of Elstner and coworkers also added a second order expansion term E_{2nd}. It is based on an interaction function γ_{ab}:

$$\gamma_{ab} = \frac{1}{r_{ab}} - S\left(r_{ab}, U_a, U_b\right) \tag{1.5}$$

which takes chemical hardness (i.e., the Hubbard parameters U_a, U_b) into account, allowing for delocalization and charge transfer between different atoms. $\gamma_{\alpha\beta}$ is based on the interaction of free atomic charges that are truncated after the monopole term during multipole-expansion resulting in spherically shaped densities and restrains the interaction between atom types a and b according to their reciprocal equilibrium distance r_{ab} and their chemical hardness. With γ_{ab}, the second order expansion term is given by $E_{2nd} = \frac{1}{2} \sum_{a,b} \Delta q_\alpha \Delta q_\beta \gamma_{ab}$. It accounts for the mutual induction of the charges Δq_a and Δq_b, where $\Delta q_a = q_a - q_a^0$.

In the implementation of DFTB3, the interaction functional γ_{ab} was revisited by making the Hubbard parameter (chemical hardness U_a) charge dependent, which makes it possible to have different chemical hardness parameters for anions, neutral molecules, and cations [44]. The extension of E_{2nd} to E_{3rd} is mediated by the set of Equations 1.5:

$$\begin{aligned}
\Gamma_{ab} &= \left.\frac{\partial \gamma_{ab}}{\partial q_a}\right|_{q_a^0} = \left.\frac{\partial \gamma_{ab}}{\partial U_a}\frac{\partial U_a}{\partial q_a}\right|_{q_a^0} \quad \text{with } a \neq b \\
\Gamma_{ba} &= \left.\frac{\partial \gamma_{ab}}{\partial q_b}\right|_{q_b^0} = \left.\frac{\partial \gamma_{ab}}{\partial U_b}\frac{\partial U_b}{\partial q_b}\right|_{q_b^0} \quad \text{with } a \neq b \\
\Gamma_{aa} &= \left.\frac{\partial \gamma_{aa}}{\partial q_a}\right|_{q_a^0} = \left.\frac{1}{2}\frac{\partial \gamma_{aa}}{\partial U_a}\frac{\partial U_a}{\partial q_a}\right|_{q_a^0}
\end{aligned} \tag{1.6}$$

Using the definition of Γ (Equation 1.6) and applying analogous approximations as for the second order integrals, one finds $E_{3rd} \approx E^\Gamma = \frac{1}{3} \sum_{a,b} \Delta q_a^2 \Delta q_b \Gamma_{ab}$ [44]. Thus, Equation 1.2 is the combination of the original DFTB expression, Equation 1.3, and the second (E_{2nd}) and third order expansion terms (E_{3rd}). The matrix elements of

the full Hamiltonian of DFTB3 are given by

$$H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu\nu} \sum_c \Delta q_c \Big(\frac{1}{2}(\gamma_{ac} + \gamma_{bc}) + \frac{1}{3}(\Delta q_a \Gamma_{ac} + \Delta q_b \Gamma_{bc}) + \frac{\Delta q_c}{6}(\Gamma_{ca} + \Gamma_{cb}) \Big) \quad \forall a, b, \mu \in a, \nu \in b. \tag{1.7}$$

What is missing from the overview just given is the connection between the coefficients $c_{\nu i}$ and the charge fluctuations Δq_a. As shown in Ref. [41], this can be achieved by Mulliken charge analysis[53, 54, 55]:

$$q_a = \frac{1}{2} \sum_i^{occ} n_i \sum_{\mu \in a} \sum_\nu^N \left(c^*_{\mu i} c_{\nu i} S_{\mu\nu} + c^*_{\nu i} c_{\mu i} S_{\nu\mu} \right). \tag{1.8}$$

1.2.3 Multiscale Modeling (SQM/MM)

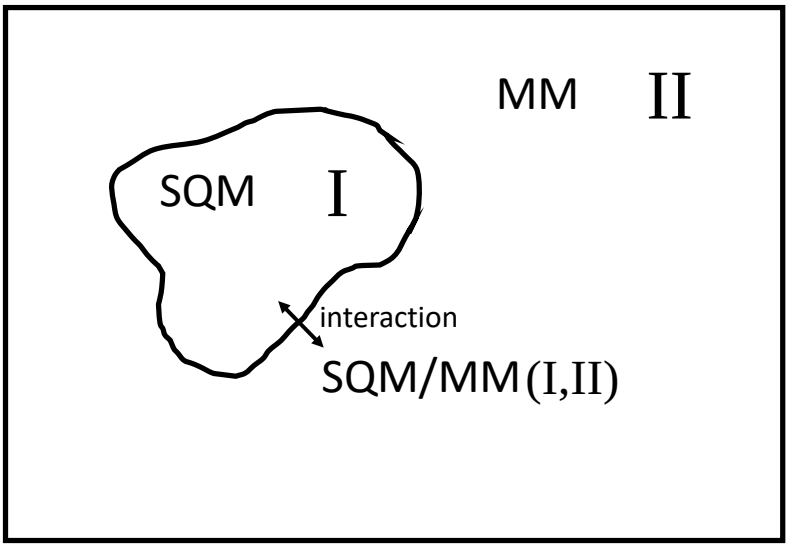


Figure 1.4: Regions of interest when coupling SQM and MM to a SQM/MM system.

Multiscale modeling generally refers to the combination of different theoretical models within a single simulation. The methodology, as already mentioned in the Introduction, which was rewarded with the Nobel Prize in chemistry, combines the accuracy of quantum mechanics with the speed of molecular mechanics and was originally published by Karplus and coworkers [56]. The combination of the theoretical models introduced in Sections 1.2.1 and 1.2.2 results in an implementation of both theoretical models (SQM/MM), which is locally divided into two regions.

Figure 1.4 illustrates the situation: Region I consists of the sub-system described with SQM, region II consists of the sub-system described with MM. This inevitably raises the question of how to treat the interactions between the sub-systems. Therefore, one has to account for the interactions between the two regions I and II. As a result of dividing the system into two different regions, the potential energy of the full system is the sum of the potential energies of each sub-system, plus the interaction between them (I,II). Usually, the (S)QM region (I) is relatively small and only contains atoms considered relevant for the calculation, while the rest of the system is described by MM (II). There are generally two approaches to calculate the total energy of a multiscale modeled system, either a subtractive or an additive

scheme.

The subtractive scheme consists of one evaluation on the SQM level of theory (E_I^{SQM}) and two evaluations on the MM level of theory E_{I+II}^{MM} (the entire system) and E_I^{MM} (the SQM region):

$$E_{Sub}^{SQM/MM} = E_I^{SQM} + E_{I+II}^{MM} - E_I^{MM} \tag{1.9}$$

The total energy $E_{Sub}^{SQM/MM}$ is obtained as the sum of the quantum mechanical sub-system evaluated at the SQM level of theory and the whole system being described with MM minus the sub-system evaluated at the MM level of theory, hence the label "subtractive". Note, that the interaction between the sub-systems is described purely at the MM level of theory.

In contrast to Equation 1.9, the additive scheme, which we used in all our studies, requires the following evaluations[57]:

$$E_{Add}^{SQM/MM} = E_I^{SQM} + E_{II}^{MM} + E_{I,II}^{SQM/MM} \tag{1.10}$$

Here, E_I^{SQM} is the potential energy of sub-system I at the SQM level of theory, E_{II}^{MM} is the potential energy of sub-system II, and $E_{I,II}^{SQM/MM}$ is the interaction between the SQM (I) and MM (II) regions. All systems studied in this thesis are small molecules that can be completely described by SQM, for further details on the treatment of this situation we refer the reader to [15]. Special care must be taken when I and II share one or more chemical bonds, for example when the SQM region represents only a small part of a larger molecule, as it is usually the case for proteins. In this case, one has to use more elaborate techniques to handle the interface region properly [58].

In most QM/MM multiscale approaches, the van der Waals interactions between regions I and II are described by the classical mechanical FF, i.e., by Lennard-Jones interactions. The Coulomb interactions between the two regions, on the other hand, should take into account the influence exerted by the point charges of the MM FF directly on the SQM electronic density and vice versa. Such treatments are commonly referred to as electrostatic embedding[59].

The general scheme of any (S)QM/MM coupling in CHARMM is given by[43, 56]:

$$E^{SQM/MM} = \langle \Psi | \hat{H}^{SQM} + \hat{H}_{el}^{SQM/MM} | \Psi \rangle + E_{vdW}^{SQM/MM} + E^{MM} \tag{1.11}$$

Equation 1.11 is general and applies to SQM/MM as well as ab initio QM/MM. In the special case of SCC-DFTB, the Coulomb interaction between the two regions can be approximated by [43]:

$$\hat{H}_{el}^{SQM/MM} \approx \sum_{a \in MM} \sum_{b \in SQM} \frac{Q_a \Delta q_b}{|\vec{R}_a - \vec{R}_b|} \tag{1.12}$$

The fluctuating Mulliken charges of the SQM atoms interact with the MM nuclear point charges Q (c.f. previous Chapter 1.2.2 Δq). The interaction of the two regions is handled self-consistently; cf. Equation 1.4 and is linked via the Mulliken charge analysis procedure(1.8).

1.3 Indirect cycle (S)QM/MM FES

1.3.1 Challenges when computing free energy differences between levels of theory

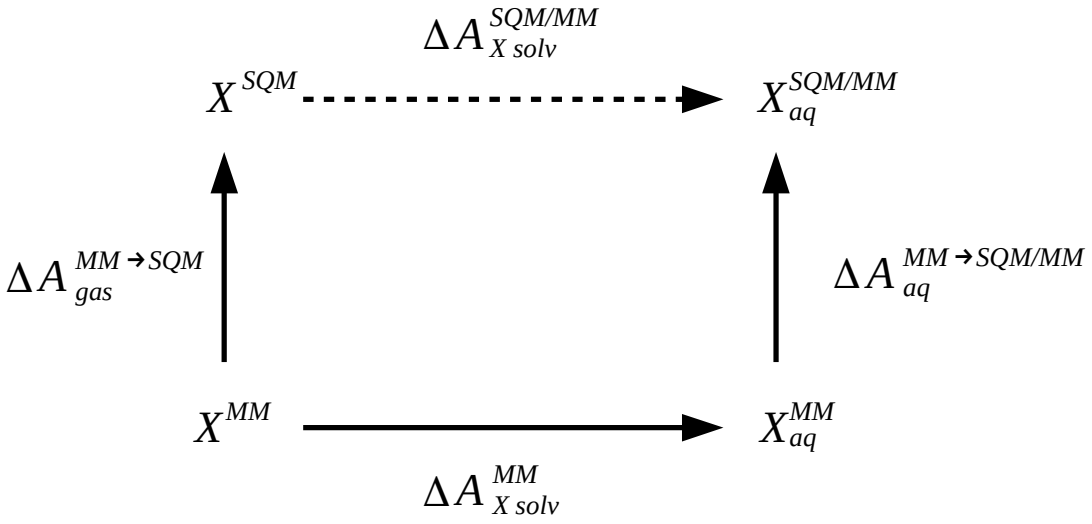


Figure 1.5: Indirect cycle to calculate the solvation free energy $\Delta A^{SQM/MM}_{X\ solv}$ of a solute X at the SQM/MM level of theory (dashed arrow).

Figure 1.5 is an enlarged version of the corrections from a low to a high level of theory outlined in 1.3. The notation has been adapted to better reflect the work done in this thesis. (1) Since we focus on a single solute, this compound is referred to as X, rather than A, B. (2) The low and high levels used are indicated explicitly, i.e., MM for the low and SQM(/MM) for the target high level of theory. (3) Since most of our work has been done at constant volume (in the NVT ensemble), we write ΔA instead of ΔG. Since the free energy is a state function, the solvation free energy $\Delta A^{SQM/MM}_{X\ solv}$ at the SQM/MM level of theory (dashed arrow in Figure 1.5) can be calculated from the other three legs of the cycle according to

$$\Delta A^{SQM/MM}_{Xsolv} = \Delta A^{MM \rightarrow SQM/MM}_{Xaq} - \Delta A^{MM \rightarrow SQM}_{Xgas} + \Delta A^{MM}_{Xsolv} \tag{1.13}$$

The advantage of this stepwise approach is that any extensive sampling can be done at the MM level of theory. The MM solvation free energy $\Delta A^{MM}_{X\ solv}$ can be computed by any of the standard force-field methods/programs, e.g., by using `transformato` [60, 61]. This is not only (much) faster, but also allows the use of the various tricks needed to perform alchemical transformations [28]. The free energy

differences corresponding to the vertical arrows, $\Delta A_{Xgas}^{MM\rightarrow SQM}$ and $\Delta A_{Xaq}^{MM\rightarrow SQM/MM}$, can be calculated, at least in principle, by Zwanzig's equation [21]

$$\Delta A^{MM\rightarrow SQM} = -k_B T \ln\left\langle \exp\left(\frac{-\Delta U^{MM\rightarrow SQM}}{k_B T}\right)\right\rangle_{MM} \tag{1.14}$$

Here, k_B and T have the usual meaning of the Boltzmann constant and (absolute) temperature. The energy differences $\Delta U^{MM\rightarrow SQM}$ are computed from configurations saved during trajectories at the MM level of theory, both in the gas phase and in aqueous solution. The energies of these coordinate frames are re-evaluated at both levels of theory. While this requires a certain number of SQM(/MM) calculations, all sampling is done at the MM level of theory.

The thermodynamic cycle shown in Figure 1.5 and its associated Equation 1.13 suggest that by following the indirect path, free energy differences can be computed efficiently at a high level of theory. However, Heimdal and Ryde[37] demonstrated that in practice it is almost impossible to obtain converged results for the corrections, $\Delta A_{Xgas}^{MM\rightarrow SQM}$ and $\Delta A_{Xaq}^{MM\rightarrow SQM/MM}$, when using Zwanzig's equation (Equation 1.13), unless a "frozen" QM region is used. Essex and coworkers[62] identified one source for the poor convergence. Most of the bond lengths and angles sampled at the low level of theory are subtly "wrong" at the high level of theory. Therefore, the resulting distribution of the energy differences $\Delta U^{MM\rightarrow SQM}$ becomes very broad, and the exponential mean in Equation 1.14 fails to converge.

Uni-directional methods, such as Zwanzig's equation, are known to converge only when the two end states are similar, see, e.g., Ref. [24]. Two-sided methods, such as Bennett's Acceptance Ratio method (BAR)[23], are known to converge even for larger differences between the two end states [63]. Adapted to the specific use case of connecting two levels of theory, the central equation of BAR is

$$\Delta A^{MM\leftrightarrow SQM} = k_B T \left(\ln \frac{\langle f(\Delta U^{SQM\rightarrow MM} + C)\rangle_{SQM}}{\langle f(\Delta U^{MM\rightarrow SQM} - C)\rangle_{MM}}\right) + C \tag{1.15}$$

In Equation 1.15, $f(x) = \frac{1}{1+e^x}$ is the Fermi function, and C a constant to be determined. In practice, C is iteratively refined until the numerator and denominator in the argument of the logarithm in Equation 1.15 become identical, and hence the logarithm vanishes. Once C has been determined in this manner, the free energy difference becomes

$$\Delta A^{MM\leftrightarrow SQM} = -k_B T \ln \frac{N_{SQM}}{N_{MM}} + C \tag{1.16}$$

Here N_{MM} and N_{SQM} denote the number of samples at the MM and SQM levels of theory, respectively.

An obvious disadvantage of using BAR is that one now needs to sample also at the high, i.e., in our case, the SQM(/MM) level of theory, as indicated by the ensemble mean $\langle\ \rangle_{SQM}$ in Equation 1.15. Successful applications of BAR have been reported [64, 65]. However, there are also several documented failures, e.g., [66, 67]. The subtle "mismatches" between bond lengths and angles at the two levels of theory (see above) make the distributions of both the forward and backward energy differences very broad, and often there is no overlap. When this is the case, BAR is not guaranteed to give correct results[23, 24]. As described earlier, in regular, force field-based FES one introduces alchemical intermediate states whenever the two endpoints are too dissimilar [24]. However, in the context of calculating free energy differences between levels of theory, this is not practical as the computational cost would be too high.

1.3.2 Nonequilibrium Work Switching Simulations (NEW)

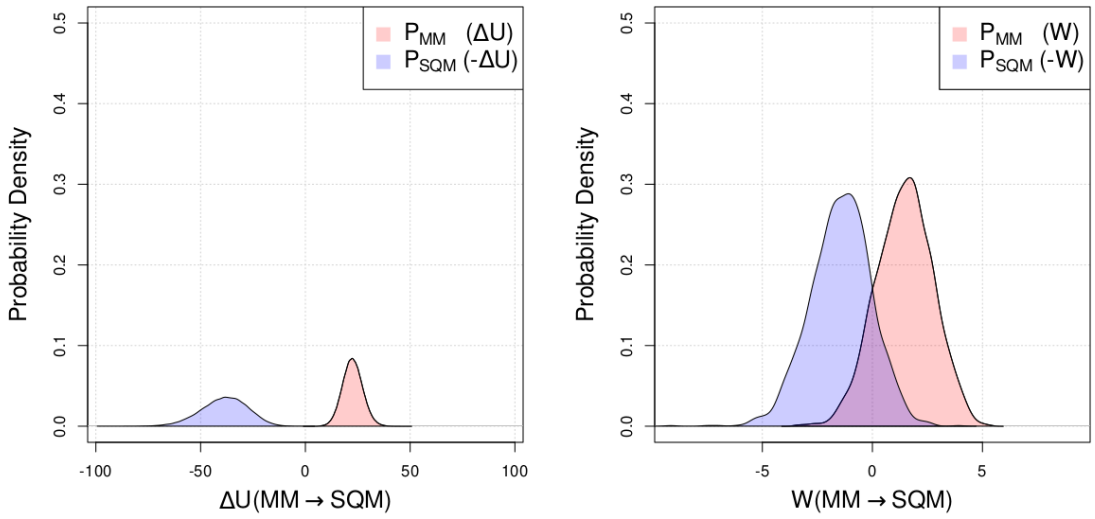


Figure 1.6: Comparison of the calculation of the free energy difference between MM and SQM via FEP(left) and via NEW(right).

As just described in Section 1.3.1, a major challenge when calculating free energy differences between levels of theory, i.e., $\Delta A_{Xgas}^{MM\rightarrow SQM}$ and $\Delta A_{Xaq}^{MM\rightarrow SQM/MM}$ in Figure 1.5, is the lack of overlap between the distributions of forward and backward energy differences. The situation for BAR is shown on the left side of Figure 1.6. Not only are the two distributions very broad, but there is virtually no overlap between the forward (red) and backward (blue) densities.

Nonequilibrium work (NEW) methods are an alternative to equilibrium techniques for calculating free energy differences. In 1997, Jarzynski showed that Zwanzig's equation can be extended to the nonequilibrium case [68].

$$\Delta A^{MM\rightarrow SQM} = -k_B T \ln\left\langle \exp\left(\frac{-W^{MM\rightarrow SQM}}{k_B T}\right)\right\rangle_{MM} \tag{1.17}$$

Equation 1.17, hereafter abbreviated as JAR, is completely analogous to Equation 1.14 (Zwanzig's equation), but the energy differences $\Delta U^{MM\rightarrow SQM}$ are replaced by the work W of switching the system from an MM to a SQM(/MM) description of the interactions. The nonequilibrium generalization of BAR was derived by Crooks[69]:

$$\Delta A^{MM\leftrightarrow SQM} = k_B T\left(\ln\frac{\langle f(W^{SQM\rightarrow MM}+C)\rangle_{SQM}}{\langle f(W^{MM\rightarrow SQM}-C)\rangle_{MM}}\right)+C \tag{1.18}$$

As in the case of JAR vs. Zwanzig's equation, Equation 1.18 is analogous to Equa-

tion 1.15; forward and backward energy differences are replaced by the corresponding work values W. The use of Equation 1.18 will henceforth be abbreviated as CRO.

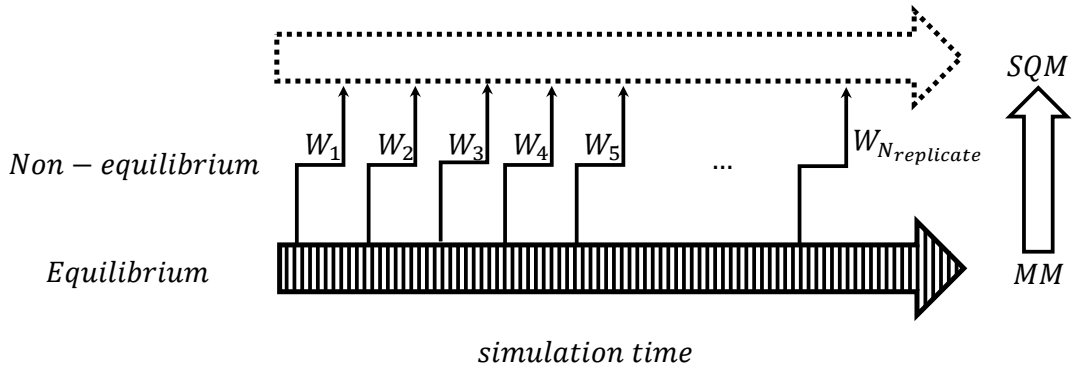


Figure 1.7: Illustration of the calculation of $\Delta A^{MM \to SQM}$ via NEW.

The workflow for computing NEW values for the transition from MM to SQM is illustrated in Figure 1.7; see also Ref. [70]. One starts with an equilibrium simulation at the MM level of theory (thick, solid arrow) and saves coordinates and velocities at regular intervals; in practice, restart files were written. Then, one selects a sufficient number of restart files, which ideally are statistically independent (i.e., not correlated) and uses them to start (very) short simulations during which the energy function is gradually changed from the MM to the SQM(/MM) level of theory. This is indicated by the vertical arrows, resulting in work values W_1 through $W_{N_{replicate}}$. The simulation length N_{switch} for each switch is represented by the small horizontal "bend" in the arrows.

The mixing of the MM and SQM(/MM) Hamiltonians is done with the MSCALE functionality of CHARMM [71]. The mixing is controlled by a time-dependent coupling parameter $\lambda_t = \lambda(t)$:

$$U(\lambda_t) = (1 - \lambda_t)U^{MM} + \lambda_t U^{SQM/MM} \tag{1.19}$$

As usual, $0 \leq \lambda_t \leq 1$ and the increment follows

$$\lambda_t = \lambda_t(i) = i/N_{switch}, \qquad i = 1, 2, \ldots, N_{switch} \tag{1.20}$$

where N_{switch} is the number of MD steps of a single switching simulation. The number of steps, N_{switch}, is typically between 200 and 10,000, i.e., for a time step of 1 fs, this corresponds to a simulation length or each switch of 200 fs to 10 ps. During each switch, the work W is accumulated with CHARMM's PERT module[5]

in slow-growth mode according to

$$W(t+\delta t) = W(t) + U(x_{t+\delta t}, \lambda_{t+\delta t}) - U(x_{t+\delta t}, \lambda_t) \tag{1.21}$$

The work values obtained from the switching simulations are the input needed for JAR, which we rewrite here to make the averaging process explicit:

$$\Delta A^{MM\to SQM} = -k_B T \ln \left(\frac{1}{N_{replicate}} \cdot \sum_{i=1}^{N_{replicate}} \exp\left(\frac{-W_i^{MM\to SQM}}{k_B T} \right) \right) \tag{1.22}$$

Comparing Equation 1.22 (or 1.17) with Equation 1.14, it is clear that using JAR is more expensive. Instead of just recomputing a few hundred (or a few thousand) energies at the SQM(/MM) level of theory, one now has to perform $N_{replicate}$ NEW switches, each consisting of N_{switch} steps. Thus, the total cost of calculating a low to high correction is $N_{replicate} \times N_{switch}$ MD steps, each of which requires the calculation of energies and forces at both the MM and SQM(/MM) levels of theory. It should be noted, however, that the switching simulations are independent, i.e., they can be run in parallel. To use CRO, one also needs work values in the high to low direction. In principle, these can be obtained by reversing the scheme shown in Figure 1.7. The most expensive part is the need to run sufficiently long equilibrium simulations at the high level of theory to obtain the starting points for the NEW switches from the high to the low level of theory.

The use of JAR and CRO to calculate free energy differences between levels of theory has been extensively tested in several studies[72, 73, 66, 74, 67]. The results are encouraging and prove that NEW-based approaches are reliable methods for calculating the end state corrections, $\Delta A_{gas}^{MM\to SQM}$ and $\Delta A_{aq}^{MM\to SQM/MM}$, required for Eq. 1.13. In these studies, CRO was used as the gold standard to validate the results obtained with JAR. However, even with JAR, the protocols used were very extensive, up to several thousand picosecond switches. The focus of these initial studies was to demonstrate correctness and reliability. For most practical applications, however, these extensive protocols would be too expensive in terms of computer time. To be useful in practice, the protocols of Refs. [72, 73, 66, 74, 67] need to be optimized while maintaining their reliability.

1.4 Overview of the thesis

Therefore, the overarching goal of this work was to streamline the previously used NEW protocols so that JAR can be employed to calculate end state corrections in real-world problems. The gas phase ($\Delta A_{gas}^{MM \rightarrow SQM}$) and solution steps ($\Delta A_{aq}^{MM \rightarrow SQM/MM}$) were optimized separately; this work is presented in Chapter 2 and Chapter 3, respectively. Since the results of these chapters have been published in peer-reviewed journals, Chapters 2 and 3 are concise summaries of the publications. The full text of the journal articles is reproduced in Chapters I and II of Publications. The supporting information for the two articles is provided in Chapters A and B of Appendices.

The third project reported here is a first application of the optimized protocols to the calculation of absolute solvation free energies, combining NEW switching simulations with the existing `transformato` framework[60, 61] for MM-based free energy simulations. The results are described in Chapter 4. The thesis concludes with a discussion of the results and problems encountered, as well as an outlook on future work (Chapter 5).

2. Summary of Publication I – Optimizing the gas phase

Note: The full text of the manuscript summarized here can be found in Chapter I of Part Publications.

During the last decade, the accurate calculation of free energy differences between levels of theory has been the subject of intense research [75, 29, 76, 77, 78, 79, 64]. Approaches based on nonequilibrium work (NEW) switching simulations were introduced and shown to be reliable [72, 73, 66, 74, 67]. The focus of these early studies was on proving correctness; the protocols used were too expensive for practical applications. In addition, even a very short series of NEW switching simulations requires significantly more computational resources than approaches based on equilibrium methods. The high computational cost prevented the method from being used by a wider audience.

Therefore, the central goal of my thesis was to optimize the previously used NEW protocols to make them applicable to real-world problems. This chapter focuses on gas phase transformations, trying to find the optimal balance of resources while keeping the computational cost constant. As model compounds for this study, we chose a subset of the so-called HiPen data set.[67] HiPen stands for "high penalty" and refers to the penalty scores reported by CGenFF when FF parameters are assigned by analogy [80, 81, 82]. All molecules in the set have such high penalties, so calculating the free energy difference between levels of theory is expected to be challenging. Since we want to optimize protocols, not to solve convergence problems, we chose mostly the so-called "good" compounds, i.e., those molecules for which converged results were obtained with JAR (See Figure I.1 of Publication I). As we try to avoid sampling at the high level of theory, our focus was primarily on the forward switching direction (from MM to SQM). As in previous work, bidirectional calculations employing CRO were only used to obtain reference results.

Recalling the workflow how to use NEW switching simulations and JAR to calculate a free energy difference (cf. 1.3.2) leads to an immediate question concerning resource allocation: One can either carry out many short switches, or fewer, correspondingly longer switches. This is illustrated in Figure 2.1. To answer this question, we evaluated several combinations of switching length (N_{switch}) and number of switches ($N_{replicate}$) that result in the same cumulative simulation length (t_{total}) and thus the same computational cost when computed on a single CPU. For example, 8,000 200 fs switches result in the same t_{total} = 1.6 ns as 1,600 1,000 fs switches.

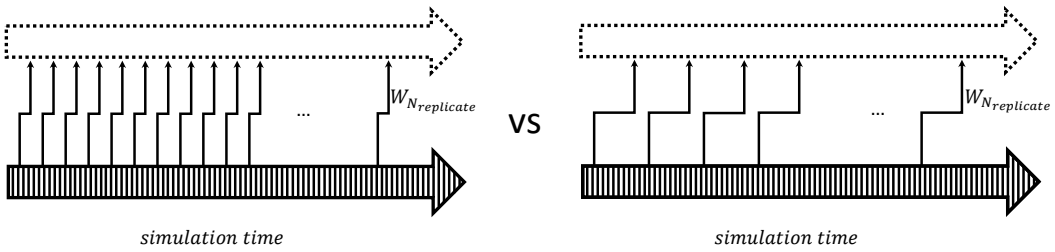


Figure 2.1: The main question addressed by Publication [83]: Is it better to use more, shorter (left) or fewer, longer switches (right)?

Specifically, we compared four combinations, namely $8,000 \times 200$ fs, $3,200 \times 500$ fs, $1,600 \times 1,000$ fs, and $800 \times 2,000$ fs switches. Analyzing the initial set of results, we noted unexpected convergence problems for some compounds. To investigate these, we developed several tools to trace the origin of this problem, which will be described in more detail below. Specifically, we discovered that in some cases the starting points for the NEW switches which were not sufficiently equilibrated. As expected, the results obtained repeating the full series of calculations were much more consistent.

	[kcal/mol]	
	MAD	RMSE
NSWI200	0.27	0.31
NSWI500	0.08	0.09
NSWI1000	0.08	0.09
NSWI2000	0.04	0.05
NSWI2000$^{\text{red}}$	0.06	0.08

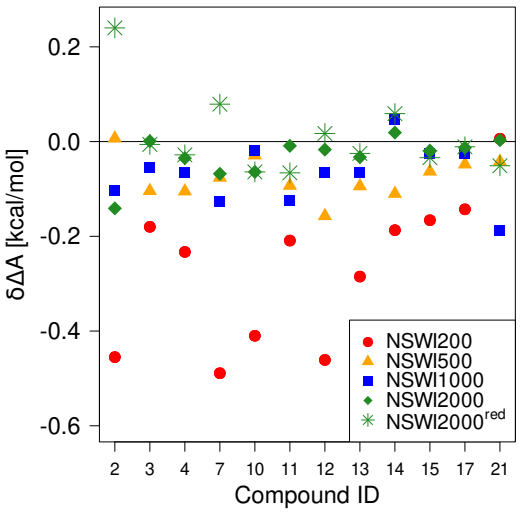


Figure 2.2: Final results obtained in Publication I [83]: MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) of results obtained from forward switches and Jarzynski's equation compared to Reference result (CRO).

The corrected free energy differences also clearly answer the central question: The longer the switching length, the better the results. Thus, the NSWI2000 protocol using $800 \times 2,000$ fs switches gave by far the best overall result, i.e., had the lowest

deviation from the CRO reference results; cf. Figure 2.2. The corresponding lower number of switches seemed to have no effect on the quality and convergence of the results. We, therefore, reduced the number of 2,000 fs switches from 800 to 200. While there was some deterioration, the quality of this reduced protocol (NSWI2000$^{\text{red}}$) was still better than that of the four times more expensive $1,600 \times 1,000$ fs one (NSWI1000). When we attempted to reduce the number of switches even further, the deviations from the CRO reference results increased considerably, at least for some test systems.

To trace the convergence problems mentioned above, we developed and used two novel ways to analyze the raw data. When free energy differences are computed by JAR or CRO, the ordering of the work values is irrelevant. Nevertheless, the starting points for the switching simulations are generated in a deterministic, temporal order. Specifically, in this work, the snapshots were saved during eight MD simulations initialized with different random velocities. Plotting the work values according to the temporal order of their starting points, and concatenating the eight data sets from the independent MD simulations as if they were one long simulation, not only makes it easy to identify "outliers," but also where in the sequence of switches they occur. For an illustrative example, see Figure 2.3. We use the word "outlier" loosely to refer to work values which do not lie within three standard deviations of the mean $\bar{W}$. Looking at the work values of the initial set of results in this manner, which we refer to as "quasi time series analysis", revealed that most of the unusual work values clustered at the beginning of the eight individual simulations. With this information, it was easy to find the source of the problem, an insufficient amount of equilibration.

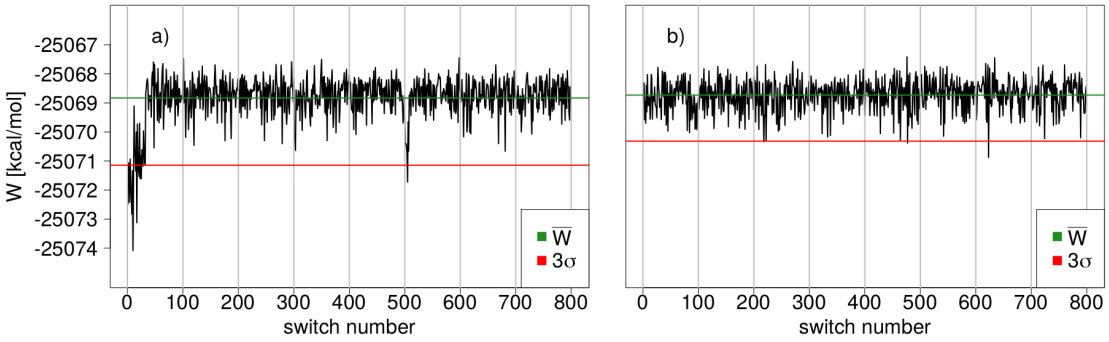


Figure 2.3: Illustrative example of a use case of "Quasi time series" analysis for compound **21** of Publication I[83]. The first and the sixth bin of a) clearly shows a suspect behavior at the beginning of the respective bin. The corrected protocol (longer equilibration) led to the work values shown in b).

Applying our quasi time series analysis to the corrected second set of results,

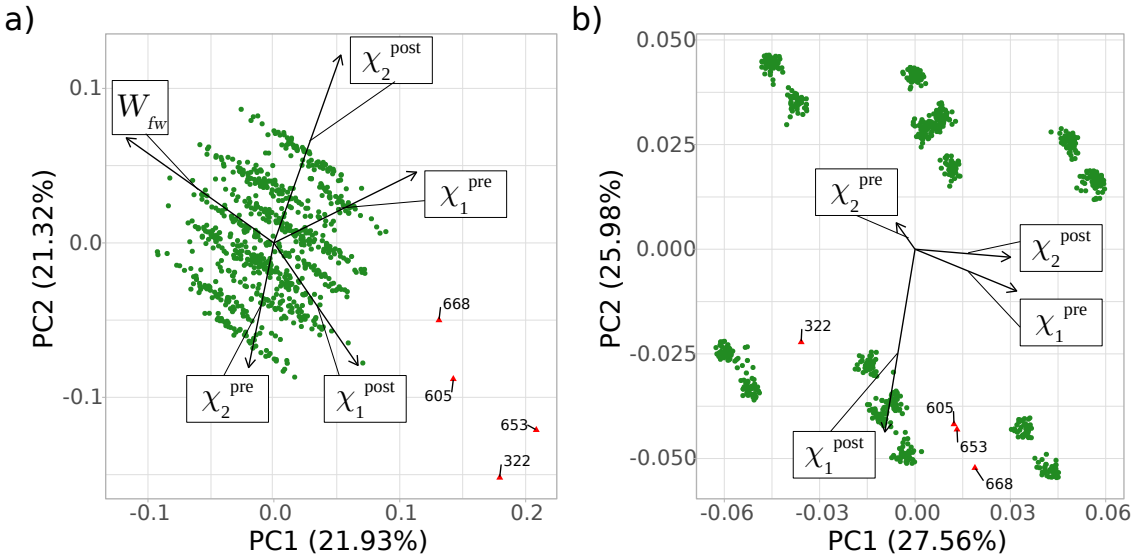


Figure 2.4: Illustrative example of a use case of PCA analysis to connect negative outliers to the dihedral angles of compound **2** from Publication I[83]. a) shows a scaled PCA including the work values. b) shows an unscaled PCA, including only the dihedral angles prior and after the switch. Outliers are identified automatically and are the same in both cases.

we still found occasional work values that deviated from $\bar{W}$ by more than three standard deviations toward more negative values. Such rare, negative work values can lead to systematic deviations of the free energy differences calculated by JAR, since they receive a high weight in the exponential averaging (cf. Equation 1.6). To associate these unusual work values with the sampled conformations, we performed a Principal Component Analysis (PCA) using the work values and two relevant dihedral angles before and after the switch raw data. Automatic color coding of outliers in the resulting plots made it possible to associate highly negative work values with unusual behavior of these dihedral degrees of freedom. For an illustrative example, see Figure 2.4. For two systems studied in detail, we found that the negative outliers were caused by a discrepancy in conformational sampling between MM and SQM. Specifically, we were able to show that the terminal dihedral angle of an oxime group, which involves the rotation of a hydrogen atom into a hydroxyl group, is sampled exclusively in the trans configuration in MM, but can occur in both cis and trans configurations in SQM.

3. Summary of Publication II – Optimizing the solvated phase

The NSWI2000$^{\text{red}}$ protocol of Chapter 2 (Publication I) should be sufficiently efficient for real-world application, but it was tested in the gas phase only. In earlier studies, there were clear indications that convergence in aqueous solution may be more difficult to achieve [66]. We, therefore, tested the optimized protocol on eight pairs of tautomers without freely rotatable dihedral angles. We deliberately chose rigid solutes to explicitly exclude sampling problems resulting from different conformational preferences between the MM and SQM/MM levels of theory. In addition, we also included alanine and serine dipeptide, which had been used as model compounds in Ref. [66] (see Figure II.10 of Publication II). Specifically, we studied the convergence of the free energy difference between a full MM description of the system (all molecules solvated in a water box with side-length of ≈ 26 Å and the SQM/MM representation of the system. At the target high-level of theory, the solute was described by SCC-DFTB, expanded to third order [44] using the 3ob-3-1 parameters [45, 46, 47, 48], whereas solvent waters continued to be modeled by the TIP3P model [84]. For the full description of the computational details, see Section II.4 of Publication II.

When using the NSWI2000$^{\text{red}}$ protocol (the results labeled MM=JAR(2ps) in Figure 3.2), the $\Delta A^{MM\rightarrow SQM/MM}_{X\ solv}$ correction free energy deviated significantly (> 1 kcal/mol) from the CRO$_{REF}$ results for two of the model compounds (5-t2, 6-t2; these are indicated by red arrows). Even using switching lengths of 5 ps (MM=JAR(5ps)) rather than 2 ps did not lead to satisfactory results for one solute (6-t2); the deviation to the reference result remained high (red circle in Figure 3.2). When using electrostatic embedding, Lennard-Jones interactions between the solute and the solvent are identical at the two levels of theory; the only relevant change is the charge distribution of the solute (the SQM region). We, therefore, explored two approaches to facilitate convergence of $\Delta A^{MM\rightarrow SQM/MM}_{X\ solv}$ while keeping the computational cost low. First, similarly to the work by Ito and Cui [65], we inserted an intermediate state, still completely described at the force-field level, with partial charges resembling the charge distribution at the high level of theory. Second, instead of switching linearly from the low to the high-level of theory, we attempted to use stepwise linear paths with two or three switching rates. Despite testing numerous protocols, we did not find one which gave consistent improvements in all cases. For the detailed description of these ultimately fruitless attempts, see Section II.2.4 of Publication II.

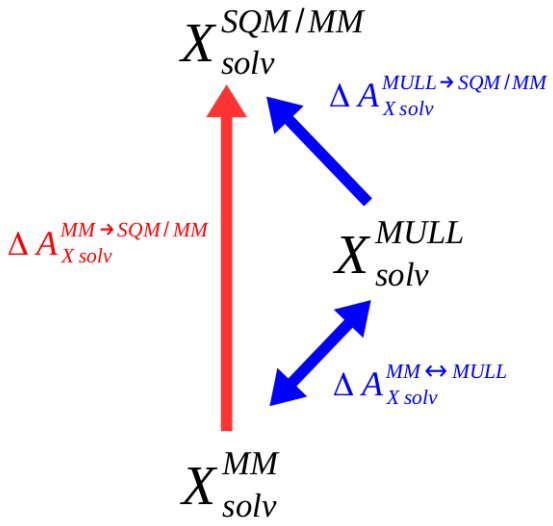


Figure 3.1: Introduction of a hybrid charge intermediate state (MULL) to reduce convergence problems when computing the free energy difference between MM and SQM/MM levels of theory.

The central idea of introducing a hybrid charge intermediate state is shown in Figure 3.1. Since SCC-DFTB uses Mulliken charges internally, average Mulliken charges were used in this work. If different quantum chemical methods are used as the high level of theory, other charge assignment schemes could (and should) be explored. We tested three ways how to obtain a representative sample of configurations for which average Mulliken charges were computed. (i) Clearly, the most exact method is to use simulations at the high level of theory (the solute treated by SCC-DFTB, surrounded by MM waters) to obtain configurations from which the Mulliken charges are calculated and then averaged. Results using this intermediate state are labeled *MULL(solv*)*. Our goal, however, is to avoid extended simulations at the high level of theory as much as possible. Therefore, we also used configurations sampled at the MM level of theory to derive Mulliken charges for the intermediate representation. Specifically, we used (ii) the Mulliken charges obtained from snapshots of the MM gas phase simulation of the solute (labeled as *MULL(gas)*), and (iii) snapshots from the corresponding MM simulations in water to obtain Mulliken charges for the solute (labeled as *MULL(solv)*). When using an intermediate state as illustrated in Figure 3.1, one also has to compute the free energy difference $\Delta A^{MM \leftrightarrow MULL}_{X\ solv}$; however, this is a free energy calculation at the MM level, the computational cost of which is small. For the full details of the averaging procedure, see Section II.4.3 of Publication II.

	[kcal/mol]	
	MAD	RMSE
MM=JAR(2ps)	0.29	0.15
MM=JAR(5ps)	0.12	0.10
MULL(gas)	0.12	0.19
MULL(solv)	0.09	0.14
MULL(solv*)	0.07	0.06

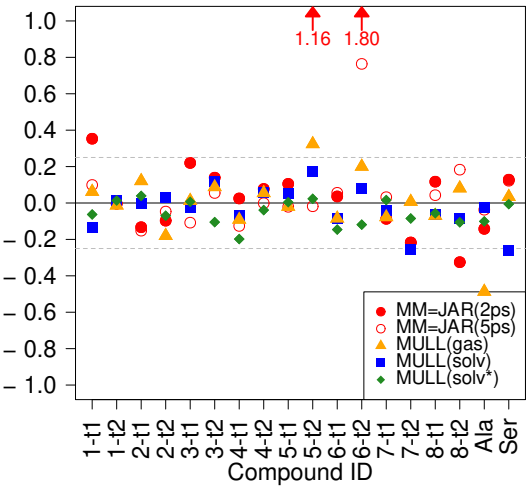


Figure 3.2: Summary of main results obtained in Publication II: MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) of results obtained from forward switches and Jarzynski's equation compared to Reference result (CRO(5ps)). All results obtained with hybrid charge intermediates (MULL(gas), MULL(solv) and MULL(solv*)) used a switching length of 2 ps.

The results obtained with the three hybrid intermediate charge states are shown in Figure 3.2 and contrasted with the results of the direct 2 and 5 ps switches. One sees that the use of either of the three MULL intermediates leads to better results than using 5 ps switching lengths without intermediate state (MM=JAR(5ps) results in Figure 3.2, red circles). All $\Delta A_{X\ solv}^{MULL \rightarrow SQM/MM}$ results were computed with the efficient NSWI2000$^{\text{red}}$ protocol (200 switches of 2 ps length). Not surprisingly, the *MULL(solv*)* intermediate state gave overall the best results. However, the performance of the *MULL(solv)* intermediate state, for which no sampling at the high level of theory is required, was also adequate. When using this intermediate state, which has, in fact, no computational overhead, all results are within ±0.25 kcal/mol of the respective CRO$_{REF}$ value. By introducing a suitable intermediate charge state, we could make the optimized gas phase protocol NSWI2000$^{\text{red}}$ also work for solute-solvent systems. Thus, we have a validated, efficient protocol to obtain the end point corrections needed for indirect (S)QM/MM FES in the gas phase *and* in solution, paving the way for real-world applications.

In addition to finding a reliable protocol to calculate $\Delta A_{X\ solv}^{MM \rightarrow SQM/MM}$ accurately at acceptable computational cost, we carried out detailed analyses to understand the source of the slow convergence of direct computations without intermediate state. Obviously, the charge representation of the solute by fixed partial charges will differ

Table 3.1: MAD (Mean Absolute Deviation) of all compounds for $RMSD_q$, the magnitude of the differential dipole moment $\Delta\mu$, the angle Θ_{SQM} between the dipole moment vector of the respective charge distribution and that of the MULL(solv*) charge distribution, and ΔN_{Waters}, the difference in average number of waters in the first solvation shell compared to a SQM/MM trajectory.

Pathway	MAD $RMSD_q[e]$	MAD $\Delta\mu$ [D]	MAD $\theta_{SQM}[^\circ]$	MAD ΔN_{Waters} [a)]
MM	0.14	3.38	29.8	0.90
MULL(gas)	0.06	2.44	7.6	1.14
MULL(solv)	0.04	1.81	16.0	0.47
MULL(solv*)	/	/	/	0.14

[a)] MAD ΔN_{Waters} was not reported in the original publication but was calculated from the absolute values shown in Figure II.6 of Publication II.

from that of the fluctuating charges at the SQM/MM level of theory. This suggests that the three sets of averaged Mulliken-like charges we have used as the intermediate states are more likely to resemble the SQM/MM charge distribution than the original FF charges. To support this assumption, we computed and analyzed three partial charge related quantities for all our model compounds relative to our best model, the MULL(solv*) charges. Specifically, we looked at a "root-mean-square-deviation of charges" ($RMSD_q$), the difference in magnitude of the dipole moments $\Delta\mu$, and the difference in angles between the dipole moments of the two charge distributions θ_{SQM}. The mean average deviations of these three characteristics of the MM, MULL(gas), and MULL(solv) charges using MULL(solv*) as the reference charge distribution are summarized in Table 3.1. The MAD of both the $RMSD_q$ and $\Delta\mu$ decreases in the order MM, MULL(gas), and MULL(solv), in line with the overall performance of the approaches (cf. Figure 3.2). The smallest angle deviation between dipole moments (MAD θ_{SQM}) was obtained for MULL(gas). This suggests that the orientation of the dipole moment resulting from a charge distribution is less important than its magnitude.

The obvious next question was how the differences in charge distribution of the solute affect the surrounding solvent. First, we calculated the average number of water molecules in the first solvation shell around each solute, counting all water molecules < 3 Å away from the solute for the MM and all three MULL charge representations. Here, we could directly compare with the average number of waters found in the high-level SQM/MM simulations. As for the solute properties, we compared the MAD of the number of water molecules for all model compounds. This data was not given in the original publication but is easily calculated from the absolute values listed in Figure II.6 of Publication II. Not surprisingly, the MULL(solv*) simulations

behaved very similarly to the high-level SQM/MM simulations, with a MAD in the average number of waters in the first shell of only 0.14. Depending on the functional group present in the molecule, when using MM charges the first solvation shell of the solutes contained on average one water too many or too few; the MAD was 0.90. Regarding this criterion, the MULL(gas) charge distribution performed worst, the MAD was 1.14. The MAD for the MULL(solv) charge distribution, on the other hand, was 0.47, a considerable improvement over the MM value.

Our second analysis probing solute-solvent interactions was motivated by solvation spectroscopy; we wanted to determine the average time it takes for the water molecules to reorient in response to the change in solute charge. We ran a series of 10 ps simulations starting from configurations equilibrated with one of the charge representations we studied (MM or MULL(solv)) and instantaneously changing the description of the solute to the SQM/MM level of theory. At each time step, we recorded the energy difference $\Delta U(t) = U^{SQM/MM}(t) - U^{method}(t)$, where the *method* was either MM or MULL(solv), assuming that $\Delta U(t)$ captures the energetic cost of the reorientation process of water. As $t \to 10$ ps, the individual energy differences approach a constant value (ΔU_{inf}). Averaging over all 800 such simulations and subtracting the average ΔU_{inf} resulted in a decay of the energy difference, which we fit to a single exponential function. For both the MM and MULL(solv) starting points, we found that TIP3P waters reoriented to the change in charge distribution of the solute with a relaxation time constant of approximately 1 ps. This value is in accord with experimental data from solvation spectroscopy [85, 86]. While the relaxation constants are very similar in both cases, the time $\Delta U(t)$ needs to approach ΔU_{inf} depends strongly on $\Delta U(t = 0)$, i.e., the difference in charge distributions between the initial and final state. If $\Delta U(t = 0)$ is large, as is the case for FF charges (MM) vs. the SQM/MM representation of the system, then the complete relaxation of the solvent requires at least 5 ps. If, on the other hand, the starting point is suitably modified charges which resemble the end state, e.g., MULL(solv), then 2 ps switches are sufficient. This analysis confirms conjectures made previously [66]. It makes clear why attempts to compute $\Delta A^{MM \to SQM/MM}$ directly require switching lengths of at least 5 ps, whereas suitable hybrid charge intermediates, such as MULL(solv), reduce this time to 2 ps. The resulting saving more than compensates the computational overhead of having to extract and average Mulliken charges and of having to calculate $\Delta A^{MM \leftrightarrow MULL}$.

4. Application of end state corrections to the calculation of absolute solvation free energies

4.1 Introduction

Indirect schemes are an elegant approach for the computation of alchemical free energy differences using quantum mechanical molecular mechanical (QM/MM) Hamiltonians [28, 15]; these are also referred to as end state corrections or bookending schemes [64, 87, 88]. The preceding two chapters were concerned with developing optimized protocols to compute the needed free energy differences between the low level of theory, e.g., MM, and the target high level of theory, e.g., QM/MM [83, 89]. As in the other chapters of this thesis, MM indicates the use of a classical mechanical force field. The abbreviation QM refers to ab initio quantum chemical methods, SQM to semi-empirical quantum chemical methods, and (S)QM indicates the use of either. Here we present the initial results of applying the complete indirect scheme to the calculation of absolute solvation free energy differences (ASFE).

Recently, `transformato`, an open-source tool written in Python to help set up large-scale calculations of relative binding and solvation free energy differences [60, 61], was extended to the calculation of ASFEs [30]. Therefore, we can compute ASFEs at the force field (MM) level quickly and reliably. Here, we investigate whether end state corrections to an SQM/MM level of theory can improve the agreement with experiment. The thermodynamic cycles used for this purpose are shown in Figure 4.1. The lower half of the plot (black rectangle) pertains to `transformato`. Rather than computing the ASFE, $\Delta A_{X,solv}$, directly, one computes the two free energy differences along the two black, vertical arrows. All nonbonded interactions of the solute X are turned off in the gas phase ($\Delta A^{MM}_{X \to X_0,gas}$) and in solution ($\Delta A^{MM}_{X \to X_0,aq}$), respectively. The ASFE is obtained from these two annihilation steps according to $\Delta A_{X,solv} = \Delta A^{MM}_{X \to X_0,gas} - \Delta A^{MM}_{X \to X_0,aq}$.

The green and red arrows in Figure 4.1 indicate the end state corrections in the gas phase ($\Delta A^{MM \to SQM}_{X,gas}$) and aqueous solution ($\Delta A^{MM \to SQM}_{X,aq}$). Their efficient calculation by nonequilibrium work (NEW) switching simulations and Jarzynski's equation (JAR) [68] was the subject of Chapters 2 [83] and 3 [89], respectively. The ASFE at the SQM/MM level of theory, $\Delta A^{SQM/MM}_{X,solv}$ can be calculated according to

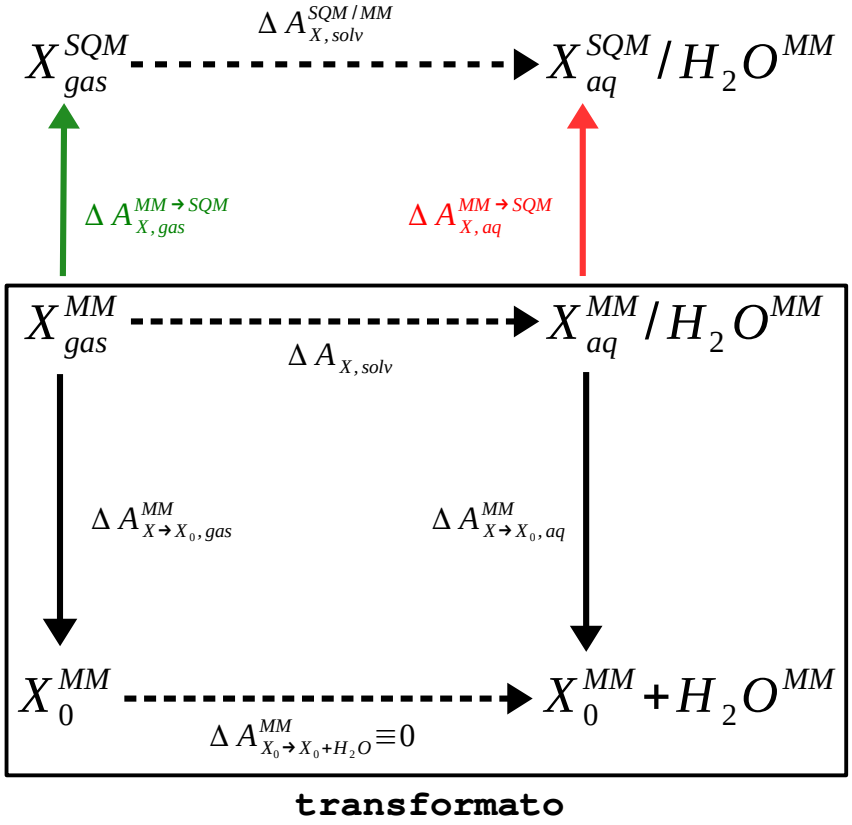


Figure 4.1: Indirect scheme used to refine absolute solvation free energies $\Delta A_{X,solv}$ calculated at the force field level using `transformato` [60, 61] to an SQM/MM level of theory.

the indirect scheme as

$$\Delta A^{SQM/MM}_{X,solv} = -\Delta A^{MM\to SQM}_{X,gas} + \Delta A_{X,solv} + \Delta A^{SQM/MM}_{X,aq}. \tag{4.1}$$

The systematic comparison of computed ASFEs to experimental data as available, e.g., in the FreeSolv database [90, 91] can be used to investigate the quality of small molecule MM force fields [30]. Here, we focus on the calculation of the ASFE at the SQM/MM level of theory for two groups of molecules. The first one (Set 1) comprises small linear and cyclic alkanes. Until today, the calculation of ASFEs for this compound class remains challenging [92, 93]. The phenomenon of better solubility of cyclic alkanes compared to normal alkanes is still not fully understood [94]. ASFEs obtained with force fields barely reflect this trend [90, 30].

The second set of molecules (Set 2) was motivated by the results obtained in the recent systematic calculation of ASFEs for the full FreeSolv database with

the CHARMM general force field (CGenFF) [80, 81, 82] using `transformato` [30]. We selected compounds for which the agreement with experiment was particularly poor. Among them are several amines, a chemical functionality for which the CGenFF results are noticeably worse compared to results obtained with the AMBER generalized force field (GAFF)[90, 91].

4.2 Material and Methods

4.2.1 Choice of model systems

Set 1: Linear vs. cyclic alkanes The first set of compounds consists of linear alkanes and their cyclic counterparts, specifically propane to hexane; see Figure 4.2. Overall, the agreement between computed and experimental ASFEs for these compounds is acceptable, both for the GAFF (RMSE 0.76 kcal/mol) [90, 91] and the CGenFF force fields (RMSE 0.72 kcal/mol) [30]; see also Table 4.4. However, as already mentioned in the Introduction, the computed ASFEs do not fully capture the differences between the linear and cyclic form of an alkane. Experimentally, the linear form is always more hydrophobic; this is particularly pronounced for cyclopropane.

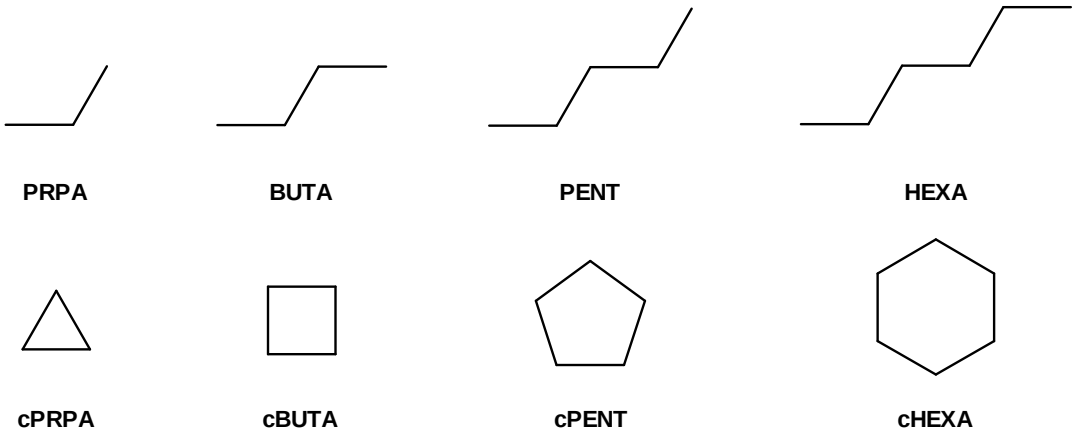


Figure 4.2: Set 1: Linear and cyclic alkanes used in this study.

Set 2: Challenging compounds from FreeSolv The second set of compounds consists of 13 molecules from the FreeSolv database[90, 91], for which the ASFEs computed with `transformato` and the CGenFF force field deviated considerably from the experimental values. As one sees in Figure 4.3, several of them are amines, a chemical functionality for which CGenFF performed poorly [30]. Other molecules with large differences between computed and experimental ASFEs were also included. In particular, the `transformato`/CGenFF result for **(13)** was considered unreliable [30] since the force field parameters suggested by the `cgenff` tool were inconsistent (identical parameters for the two sulfurs, which clearly are chemically different).

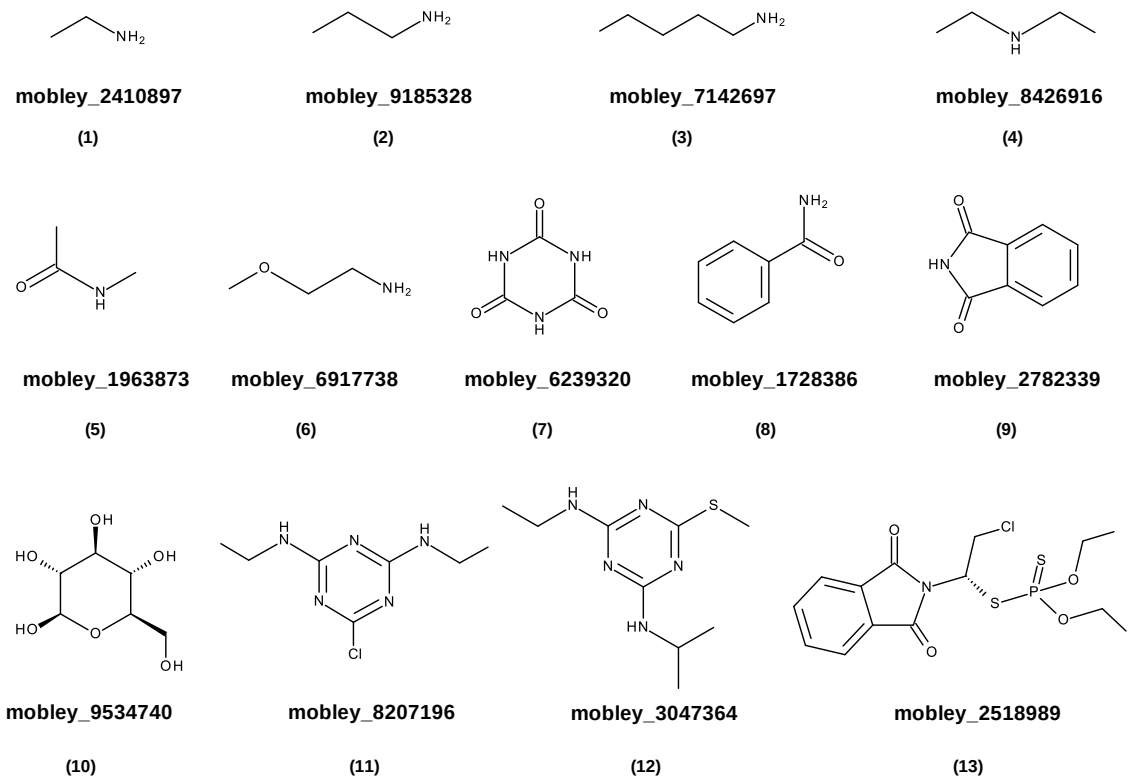


Figure 4.3: Set 2: 13 molecules from the FreeSolv database [90] for which the ASFEs computed with `transformato`/CGenFF were in poor agreement with experiment. The figure indicates the original FreeSolv identifiers as well as the simple numbering schemed used in the remainder of this chapter.

4.2.2 Simulation details

ASFEs calculated with `transformato`: For each molecule in Set 1 and Set 2, we used an automated workflow based on `macha`,[1] a standalone utility written in Python that automatizes the setup procedure of solvating molecules by using CHARMM-GUI[95, 96] generated inputs as templates, for the generation of gas phase and TIP3P solvated systems. Afterward, `transformato` was used to generate all the needed intermediate states for the calculation of the annihilation processes as described by Karwounopoulos et al.[30] Simulations for each intermediate were carried out with OpenMM(v.7.7)[97]. As the intention of this project was to apply our end state corrections to results of the regular `transformato` workflow, the ASFEs reported in Ref. [30] were used as is. However, to avoid ambiguities, we used the `transformato` raw results *without* applying the Lennard-Jones long-range correction (LRC). For further details on the calculation of absolute free energies with `transformato` we refer the reader to Karwounopoulos et al.[30]

Calculation of end state corrections by NEW simulations The following paragraphs are structured analogously to *Simulation Details* of Publication II and provide a summary overview of the simulations/calculations carried out. For additional details concerning the simulation setup, equilibration procedures, restart mechanics, etc. we refer the reader to II.4.4. The choice of protocols is based on the optimization carried out in Publication I.

Preparation and initial equilibration: For all molecules in Set 1 (Figure 4.2) except **cHEXA**, starting coordinates were generated directly by CHARMM using the internal coordinate information provided in the CGenFF topology file of model compounds. The initial coordinates were optimized by 10 steps of steepest descent and 500 steps of adopted basis Newton-Raphson (ABNR) minimization. Placement of the solutes in water boxes was carried out as described in II.4.4 under *Preparation and Initial Equilibration*. The resulting box sizes and number of TIP3P waters are summarized in Table 4.1.

All molecules in Set 2 (Figure 4.3) and **cHEXA** were solvated using the 'Solution Builder' functionality of CHARMM-GUI[95, 96]. Missing force field parameters were generated using `cgenff` version 2.5.1 [80, 81, 82] as invoked by CHARMM-GUI. The resulting box sizes and number of TIP3P waters can be found in Table 4.2. Tables 4.1 and 4.2 also list the number of atoms in each molecule, as well as the offset in kcal/mol which is subtracted from the respective end state corrections.

[1] `https://github.com/akaupang/macha`

Force field-based equilibrium simulations: Completely analogous to the workflow in II.4.4, we carried out 8 Langevin dynamics (LD) simulations (time step 1 fs, friction coefficient 5 ps^{-1}) in the gas phase and in solution. Each of these simulations was started from different random velocities. In each gas phase run, 5 ns of simulation time were discarded as equilibration, followed by 10 ns of production, during which restart files were saved every 1,000th step. In the analogous solution simulations at constant volume, 0.5 ns were discarded as equilibration. During the subsequent 1 ns production phase, restart files were saved every 1,000th step. Thus, restart files were saved at regular intervals during a cumulative simulation length of 80 ns in the gas phase and of 8 ns in solution. These served as the statistically independent starting points for the NEW switching simulations. For additional information, see *Simulation Details* of Publication II under *Force Field-Based Equilibrium Simulations*.

SQM(/MM) equilibrium simulations: To compute the end-state corrections via Crook's equation (CRO) [69], we also carried out equilibrium simulations at the SQM(/MM) level of theory. Specifically, the solute was treated with the self-consistent-charge density-functional tight-binding (SCC-DFTB) method as implemented in CHARMM [43], using the 3ob-3-1 [45, 46, 47, 48] parameter set[2]. In the solution calculations, all waters were treated classically. The cumulative simulation length in the gas phase and in aqueous solution was 8 ns. The NEW switches were started from restart files saved during these calculations. Full details can be found in II.4.4 under *SQM(/MM) Equilibrium Simulations*.

Nonequilibrium work simulations: All end-state corrections, both the unidirectional ones using JAR and the bidirectional ones using CRO, were calculated from NEW values obtained from switching simulations of 2 ps length, using a 1 fs time step. In all cases, 200 forward (MM $\rightarrow$ SQM(/MM)) and backward (SQM(/MM) $\rightarrow$MM) switches were carried out.

Calculation of $\Delta A_{Xsolv}^{\mathbf{MM \leftrightarrow MULL}}$: We only tested the calculation via hybrid charge intermediates in one case. Details for the calculation and generation of intermediate states can be found in II.4.4 under *Calculation of* $\Delta A_{Xsolv}^{MM \leftrightarrow MULL}$.

[2]https://www.dftb.org/parameters/ download/3ob/3ob-3-1-cc/

Table 4.1: Properties and simulation details of Set 1: simple alkanes and their cyclic derivatives: Compounds, their ID, the number of total and heavy atoms (N_{tot}/N_{heavy}), the number of waters used in the simulations (N_{Waters}), box length (L_{Box}) and the offset for computing free energy differences via NEW in kcal/mol.

Compound	ID	N_{tot}/N_{heavy}	N_{Waters}	L_{Box}[Å]	Offset[kcal/mol]
Propane	PRPA	11/3	572	25.64	5100
Cyclopropane	cPRPA	9/3	572	25.67	4800
Butane	BUTA	14/4	572	25.60	6600
Cyclobutane	cBUTA	12/4	572	25.61	6200
Pentane	PENT	17/5	572	25.62	8200
Cyclopentane	cPENT	15/5	572	25.64	7700
Hexane	HEXA	20/6	572	25.57	9700
Cyclohexane	cHEXA	18/6	562	26.00	9300

Table 4.2: Properties and simulation details of Set 2: challenging example compounds: Compounds, their ID, the number of total and heavy atoms (N_{tot}/N_{heavy}), the number of waters used in the simulations (N_{Waters}), box length (L_{Box}) and the offset for computing free energy differences via NEW in kcal/mol.

Compound	ID	N_{tot}/N_{heavy}	N_{Waters}	L_{Box}[Å]	Offset[kcal/mol]
mobley_2410897	**(1)**	10/3	671	28	5000
mobley_9185328	**(2)**	13/4	669	28	6000
mobley_7142697	**(3)**	19/6	668	28	9000
mobley_8426916	**(4)**	16/5	668	28	8000
mobley_1963873	**(5)**	12/5	669	28	8000
mobley_6917738	**(6)**	14/5	668	28	8000
mobley_6239320	**(7)**	12/9	665	28	14000
mobley_1728386	**(8)**	16/9	666	28	12000
mobley_2782339	**(9)**	16/11	665	28	15000
mobley_9534740	**(10)**	24/12	664	28	21000
mobley_8207196	**(11)**	25/13	837	30	19000
mobley_3047364	**(12)**	32/15	906	31	22000
mobley_2518989	**(13)**	40/23	991	32	35000

4.3 Results

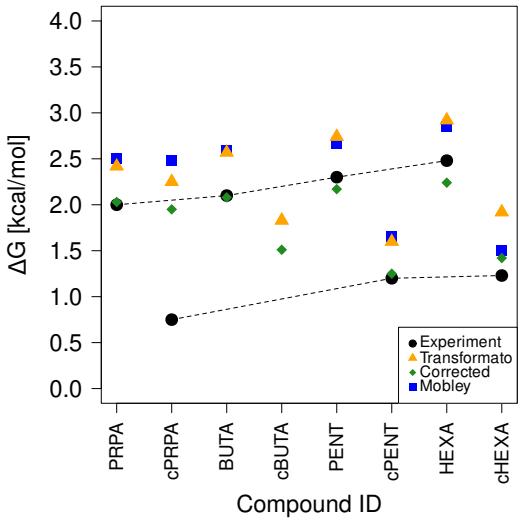
4.3.1 Set 1: Absolute solvation free energies of linear vs. cyclic alkanes

Table 4.3: Absolute solvation free energies of linear alkanes and their cyclic derivatives: All free energies are in kcal/mol.

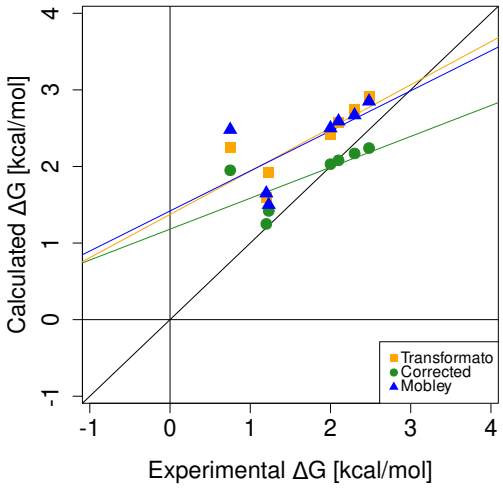
	ASFE[a)]		$\Delta A^{MM\to SQM}_{X,gas}$[b)]		$\Delta A^{MM\to SQM}_{X,aq}$[c)]		Diff[d)]	ASFE$_{corr}$[e)]	Exp[f)]	Ref[g)]
ID	ΔG	$\sigma\Delta G_i$	ΔA	$\sigma\Delta A_i$	ΔA	$\sigma\Delta A_i$	$\delta\Delta A$	ΔG	ΔG	ΔG
PRPA	2.42	0.02	-30.58	0.02	-30.97	0.02	0.01	2.03 ± 0.04	2.00	2.50
cPRPA	2.25	0.07	-46.35	0.03	-46.65	0.03	0.01	1.95 ± 0.04	0.75	2.48
BUTA	2.57	0.05	-81.95	0.03	-82.44	0.03	-0.02	2.08 ± 0.05	2.10	2.59
cBUTA	1.83	0.07	-38.64	0.05	-38.96	0.07	-0.01	1.51 ± 0.07		
PENT	2.74	0.07	-33.30	0.04	-33.87	0.04	-0.01	2.17 ± 0.07	2.30	2.67
cPENT	1.60	0.07	-59.07	0.01	-59.43	0.02	0.01	1.25 ± 0.04	1.20	1.65
HEXA	2.92	0.07	-83.93	0.03	-84.62	0.04	-0.04	2.24 ± 0.05	2.48	2.85
cHEXA	1.92	0.01	-7.08	1.26	-7.80	0.93	-1.59	1.21 ± 0.93	1.23	1.50
cHEXA*	1.92	0.01	-4.94	0.01	-5.44	0.03	-0.01	1.42 ± 0.05	1.23	1.50

[a)] Solvation free energy $\Delta A^{MM}_{X,solv}$ from `transformato` (cf. the black box in Figure 4.1) without any long-range corrections[30]; [b)] gas phase correction (green arrow of Figure 4.1), the offset applied is listed in Table 4.1;[c)] solvation correction (red arrow of Figure 4.1), the offset applied is listed in Table 4.1; [d)] Difference between solvation correction computed with JAR and CRO; [e)] solvation free energy difference $\Delta A^{SQM/MM}_{X,solv}$ at the target level of theory, computed according to Eq. 4.1; [f)] experimental value as listed in the FreeSolv database; [g)] computed values as reported in Refs. [90, 91]; *calculated with C-C-C-C dihedral angles restrained to the chair configuration.

Detailed results of the performance of `transformato`, the end state corrections in the gas and aqueous phase, the corrected ASFEs, the experimental values, and results from the Mobley group for comparison can be found in Table 4.3. Several statistical descriptors, RMSE, R^2, the Pearson, Spearman and Kendall correlation coefficients, are given in Table 4.4. Although the results obtained with `transformato` already are in good overall agreement with the experimental data (within ±1 kcal/mol), there is certainly room for improvement. We note in passing that no experimental ASFE is known for **cBUTA**. The results from `transformato` also are very similar to those reported by Mobley and co-workers using GAFF [90, 91]. ASFEs using CGenFF for **HEXA** and **cHEXA** have been reported by König et al.[98]; their values agree well with the results obtained with `transformato`. All results are also visualized in Figures 4.4a and 4.4b.



(a) Absolute values



(b) Correlation without **cBUTA**

Figure 4.4: ASFEs of alkanes and their cyclic derivatives; a) absolute values; b) correlation between computed and experimental values. Data shown includes the correction of **cHEXA***.

We assessed the convergence of the end state corrections by the deviation to the CRO_{Ref} reference result, $\delta\Delta A$, as well as the other convergence criteria used in Publications I and II, i.e., hysteresis, block standard deviation, broadness of the work distribution σW, the Π criterion suggested by Wu and Kofke [99], as well as the overlap of the forward and backward work distributions. In Table 4.3 we only report $\delta\Delta A$ for the correction in solution ($\Delta A_{X,solv}^{MM\rightarrow SQM}$). The full listing of all convergence criteria can be found under *Supplementary Material* in 4.5. We only encountered convergence problems for **cHEXA**; see below.

The results that include the end state corrections (ASFE_{corr}) are shifted towards more negative values, which, in most cases, brings the computed ASFE closer to the experimental value; see Table 4.3. The larger the solute, the larger the absolute value of the end state correction, e.g., 0.19 (**PRPA**) < 0.49 (**BUTA**) < 0.57 (**PENT**) < 0.68 kcal/mol (**HEXA**). The overall improved agreement with the experiment is reflected in the lower RMSE, 0.47 kcal/mol for the end state corrected results vs. 0.72 kcal/mol for the `transformato` results. However, even the corrected results fail to reproduce the large solvation free energy difference between **PRPA** and **cPRPA**. Therefore, the correlation between computed and experimental date is not great (see Figure 4.4b; also, the Pearson correlation coefficient for the end state corrected results is lower than for the `transformato` results).

An entirely different problem arose for **cHEXA**. As one can see in Table 4.3,

Table 4.4: Statistical descriptors for the ASFEs of simple alkanes and their cyclic derivatives: Root Mean Square Error (RMSE) in kcal/mol, Pearson correlation coefficient, Spearman correlation coefficient, Kendall rank correlation coefficient and the R^2 value of the linear regression analysis.

	RMSE [kcal/mol]	Pearson	Spearman	Kendall	R^2
`transformato`	0.72	0.80	0.89	0.81	0.64
Corrected	0.47	0.69	0.89	0.81	0.41
Mobley	0.76	0.65	0.86	0.71	0.43

this is the only system for which $\sigma\Delta A_i$ is quite high and for which we observed a sizable $\delta\Delta A$ deviation between JAR and CRO_{REF} results. A quick analysis of configurations saved during equilibrium simulations at the MM and SQM/MM levels of theory revealed that the chair conformation is adopted with different frequencies. In most of the switches, no conformational change of **cHEXA** occurs; the switching length of 2 ps is too short. In other words, almost all switches staring from the chair conformation end in the same conformation, and vice versa. The different prevalence of conformations at the two levels of theory is not reflected. We also calculated the endpoint correction for **cHEXA** restrained to the chair conformation; see the entry **cHEXA*** in Table 4.3. The uncertainty of the end state corrections is gone; similarly, $\delta\Delta A$ is also very low. We did not compute a solvation free energy difference for this restrained molecule with `transformato`; the results for **cHEXA***, therefore, are incomplete.

Table 4.5: Results for absolute solvation free energies of challenging example compounds: All results are in kcal/mol.

	ASFE[a]		$\Delta A^{MM\rightarrow SQM}_{X,gas}$[b]		$\Delta A^{MM\rightarrow SQM}_{X,aq}$[c]		Diff[d]	$ASFE_{corr}$[e]	Exp[f]	Ref[g]
ID	ΔG	$\sigma\Delta A_i$	ΔA	$\sigma\Delta A_i$	ΔA	$\sigma\Delta A_i$	$\delta\Delta A$	ΔG	ΔG	ΔG
(1)	-1.20	0.02	-301.75	0.06	-299.51	0.10	-0.03	1.04 ± 0.12	-4.50	-3.16
(2)	-1.05	0.03	-869.50	0.10	-867.22	0.13	-0.03	1.23 ± 0.25	-4.39	-3.05
(3)	-0.76	0.05	-966.64	0.14	-964.54	0.22	-0.07	1.34 ± 0.33	-4.09	-2.83
(4)	0.29	0.10	-401.66	0.72	-399.74	0.45	0.18	2.20 ± 0.64	-4.07	-2.99
(5)	-6.16	0.08	-438.68	0.04	-446.04	0.16	0.13	-13.53 ± 0.31	-10.00	-8.28
(6)	-3.15	0.08	-936.23	0.28	-934.40	0.18	-0.14	-1.32 ± 0.35	-6.55	-5.03
(7)	-29.65	0.08	-637.96	0.05	-634.04	0.33	-0.10	-25.73 ± 0.40	-18.06	-21.76
(8)	-6.09	0.02	-740.95	0.03	-748.19	0.57	-0.04	-13.33 ± 0.46	-11.00	-10.41
(9)	-18.54	0.12	-388.19	0.02	-381.27	0.83	0.20	-11.62 ± 0.59	-9.61	-11.82
(10)	-14.92	0.40	-825.15	0.46	-824.54	0.81	-0.32	-14.31 ± 0.73	-25.47	-18.09
(11)	-0.31	0.07	-671.16	0.09	-673.06	0.20	-0.10	-2.21 ± 0.33	-10.22	-10.91
(12)	-2.73	0.17	-539.12	0.27	-538.80	0.25	-0.14	-2.41 ± 0.44	-7.65	-10.55
(13)	-22.60	0.60	-98.15	0.88	-96.91	1.88	2.50	-21.36 ± 1.05	-5.74	-16.52

[a] Solvation free energy $\Delta A^{MM}_{X,solv}$ from `transformato` (cf. the black box in Figure 4.1) without any long-range corrections[30]; [b] gas phase correction (green arrow of Figure 4.1), the offset applied is listed in Table 4.2;[c] solvation correction (red arrow of Figure 4.1), the offset applied is listed in Table 4.2; [d] Difference between solvation correction computed with JAR and CRO; [e] solvation free energy difference $\Delta A^{SQM/MM}_{X,solv}$ at the target level of theory, computed according to Eq. 4.1; [f] experimental value as listed in the FreeSolv database; [g] computed values as reported in Refs. [90, 91].

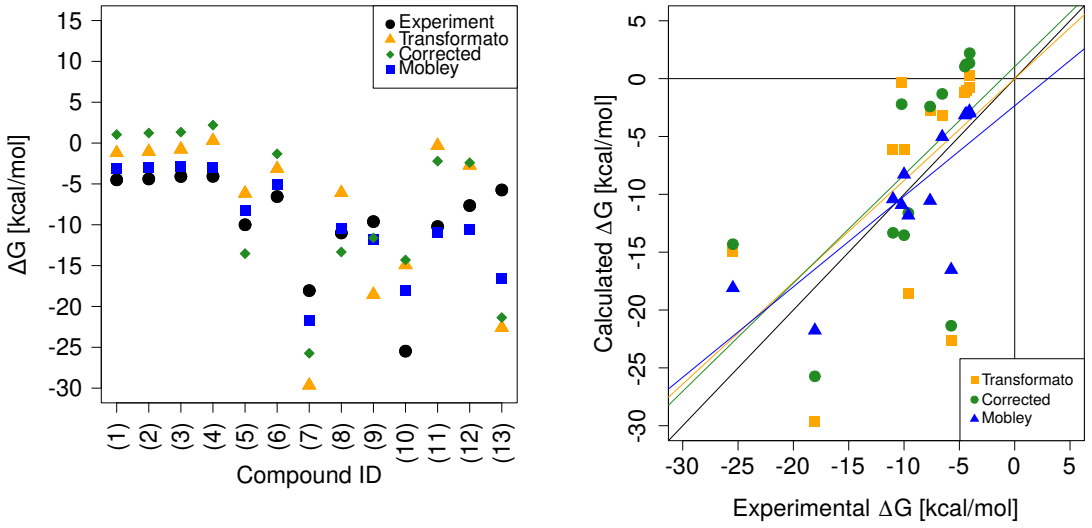
4.3.2 Set 2: Absolute solvation free energies of challenging FreeSolv compounds

Results for Set 2 are summarized in Table 4.5 and visualized in Figures 4.5a and 4.5b. The statistical descriptors of the results (RMSE, correlation coefficients, etc.) can be found in Table 4.6.

Compounds **(1)** to **(4)** are simple alkyl amines (three primary amines, one secondary amine). Experimentally, the overall trend for these four compounds is an increase in hydrophobicity with longer chain lengths. This trend is reproduced by `transformato`/CGenFF, but the computed solvation free energies are far too hydrophobic. Compounds **(5)** to **(9)** are either more complex amines or amides. The force field results obtained with `transformato` deviate in both directions from the experimental value. Compound **(7)** has the largest negative deviation to experiment (-11.59 kcal/mol). Compounds **(10)** to **(13)** are larger molecules, which are either quite flexible or contain unusual moieties. E.g., **(10)** is a carbohydrate where `transformato`/CGenFF underestimates the ASFE by $+10.55$ kcal/mol. Finally,

(13), an organophosphorodithioate, is an example of an unusual moiety. As described in Ref. [30], the force field parameters for the P-S bonds look dubious.

Looking at the quality criteria reported in Table 4.5 ($\delta\Delta A$, $\sigma\Delta A_i$), the reliability of the results for **(13)** is doubtful ($\delta\Delta A$ =2.5 kcal/mol). We, therefore, recomputed the end state correction in solution using a hybrid charge intermediate as described in *Hybrid Charge Intermediates* in Section II.4.3. While the block standard deviation $\sigma\Delta A_i$ (for the definition, see Appendix A) remained high, $\delta\Delta A$ decreased to -0.64 kcal/mol. On the one hand, this underlines the utility of using a suitable intermediate state to mitigate convergence problems; on the other hand, the resulting corrected solvation free energy is in poorer agreement with experiment. Several other compounds have somewhat larger block standard deviations $\sigma\Delta A_i$, but these seem to have little effect on $\delta\Delta A$.



(a) Absolute values

(b) Correlation

Figure 4.5: Results for absolute solvation free energies of challenging example compounds; a) absolute values; b) correlation between computed and experimental results.

The end state corrections only improved results for **(5)**, **(7)-(9)**, and **(11)**, with **(9)** being the only sizeable change in the right direction. Compounds **(7)** and **(9)** have some resemblance to the model systems of Chapter 3 as both contain a lactam functionality. Overall, applying the end state corrections decreased the RMSE from 8.01 to 7.33 kcal/mol; the other correlation metrics were improved slightly as well (see Table 4.6. This improvement, however, is mainly driven by the large correction for **(9)** from -18.54 to -11.62 kcal/mol, with the experimental value being −9.61 kcal/mol. Considering the computational effort and the remaining high

Table 4.6: Statistics for absolute solvation free energies of challenging example compounds: Root Mean Square Error in kcal/mol, Pearson correlation coefficient, Spearman correlation coefficient, Kendall rank correlation coefficient and R^2 of linear regression.

	RMSE [kcal/mol]	Pearson	Spearman	Kendall	R^2
`transformato`	8.01	0.56	0.60	0.49	0.31
Corrected	7.33	0.62	0.79	0.69	0.38
Mobley	4.02	0.78	0.69	0.64	0.61

deviation to the experimental values, end state corrections to the SCC-DFTB level of theory did not result in a significant improvement for Set 2.

4.4 Concluding Discussion

To put the results just presented, in particular those for Set 2, into perspective, one should remember that the starting point was poor. The solvation free energies of all compounds in Set 2 computed with CGenFF deviated significantly from the experimental values. Adding the end state correction to the SCC-DFTB/MM level of theory improved the overall results slightly; see the statistical descriptors (RMSE, Pearson correlation coefficient, etc.) in Table 4.6. Nevertheless, the correction pointed in the wrong direction in numerous instances. Possibly, the most disappointing or unexpected result was the failure to obtain improvements for the amines **(1)**-**(4)**. However, one should also consider that SCC-DFTB was never optimized regarding predicting ASFEs, so it may well be an inadequate high-level of theory. Finding an optimal high-level of theory for the calculation of solvation free energies is a challenging question in itself; see, e.g., Ref. [100]

An important aspect of the present results which should not be overlooked is the fact that we could compute them in a de facto routine manner. The optimizations reported in Refs. [83] and [89] were the prerequisite to make such calculations practical with acceptable computational effort. The equilibrium simulations at the MM level of theory to obtain starting structures for the switches when computed on GPUs are so fast as to be negligible (if they are not already available from the MM alchemical transformations). In solution, a 2 ps switch from the MM to the SCC-DFTB/MM level of theory requires on average 1 hour and three CPU cores. Assuming the availability of 100 cores, 200 such switches take approximately 6 hours in total.

By switching from model systems to a real application, we encountered some problems which indicate room for further enhancements. One system of interest is **cHEXA**, where there are different conformational preferences for the chair conformation at the force field and at the SCC-DFTB level of theory. The challenge here seems twofold: First, are our equilibrium simulations long enough to achieve representative sampling at the end states? Second, 2 ps switches are (far) too short so that starting from one level of theory a sufficiently large fraction of switches ends in conformations which are representative of the other level of theory, leading to the pronounced bimodal distributions of work values shown in Figure 4.6. Another, rather different example is **(13)** of Set 2. As presented in Table 4.5, the end state corrected result seems to be in slightly better agreement with the experimental data than the `transformato` result. As discussed in Results, however, this is the one system where the deviation $\delta\Delta A$ between the JAR and the CRO_{REF} results is sizable ($\delta\Delta A = 2.5$ kcal/mol); i.e., the JAR result has a systematic error. Applying

the accurate end state corrections, either the value obtained with a hybrid charge intermediate state or using the CRO_{REF} results, makes the result point slightly in the wrong direction. Thus, while the result for **(13)** of Set 2 is unsatisfactory, it contains the lesson that one has to ensure correct methodology before searching for improvements.

Errors resulting from conformational sampling may be an issue for some molecules, e.g., **13** or **cHEXA**, but seem unlikely in others, e.g., **1**–**5**. In these cases, either the SCC-DFTB level of theory is insufficient, and/or some waters must also be included in the high-level region. Frameworks for such calculations exist [35], but would require adaptation and extensive testing for such use cases. Thus, different high levels of theory, such as other SQM methods, computationally affordable QM methods, or even ML potentials, as well as the use of larger high-level regions all need to be explored in the quest for more accurately predicted free energy differences.

4.5 Supplementary Material

Table 4.7: $\Delta A^{MM\rightarrow SQM}_{X,gas}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma\Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations.

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
ID	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	$\sigma\Delta A$	W	$\delta\Delta A$
PRPA	-30.58	0.00	0.02	-30.58	0.07	2.46	2.49	30.58	0.00	0.02	30.59	0.08	2.49	2.49	-30.58	0.06	86.94	0.00
cPRPA	-46.35	0.00	0.03	-46.34	0.15	2.40	2.37	46.36	0.00	0.03	46.37	0.12	2.33	2.40	-46.36	0.06	72.08	0.00
BUTA	-81.95	0.00	0.03	-81.94	0.13	2.41	2.39	81.94	0.00	0.03	81.95	0.14	2.51	2.38	-81.94	0.06	85.36	-0.01
cBUTA	-38.64	0.00	0.05	-38.60	0.22	2.26	2.24	38.62	0.00	0.04	38.66	0.22	2.31	2.25	-38.63	0.06	79.66	-0.01
PENT	-33.30	0.00	0.04	-33.27	0.17	2.26	2.32	33.29	0.00	0.02	33.31	0.18	2.45	2.33	-33.29	0.06	83.96	-0.01
cPENT	-59.07	0.00	0.01	-59.07	0.07	2.44	2.49	59.08	0.00	0.02	59.09	0.08	2.43	2.49	-59.08	0.06	77.32	0.01
HEXA	-83.93	0.00	0.03	-83.90	0.20	2.26	2.27	83.90	0.00	0.04	83.93	0.19	2.53	2.31	-83.91	0.06	84.54	-0.02
cHEXA	-7.08	0.67	1.26	-6.24	1.27	2.61	0.93	4.93	0.00	0.02	4.97	0.34	2.61	2.24	-5.36	0.11	31.31	-1.72
cHEXA*	-4.94	0.00	0.01	-4.93	0.11	2.61	2.44	4.91	0.00	0.02	4.92	0.11	2.61	2.42	-4.39	0.06	92.82	-0.01

Table 4.8: $\Delta A^{MM\rightarrow SQM}_{X,solv}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma\Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations.

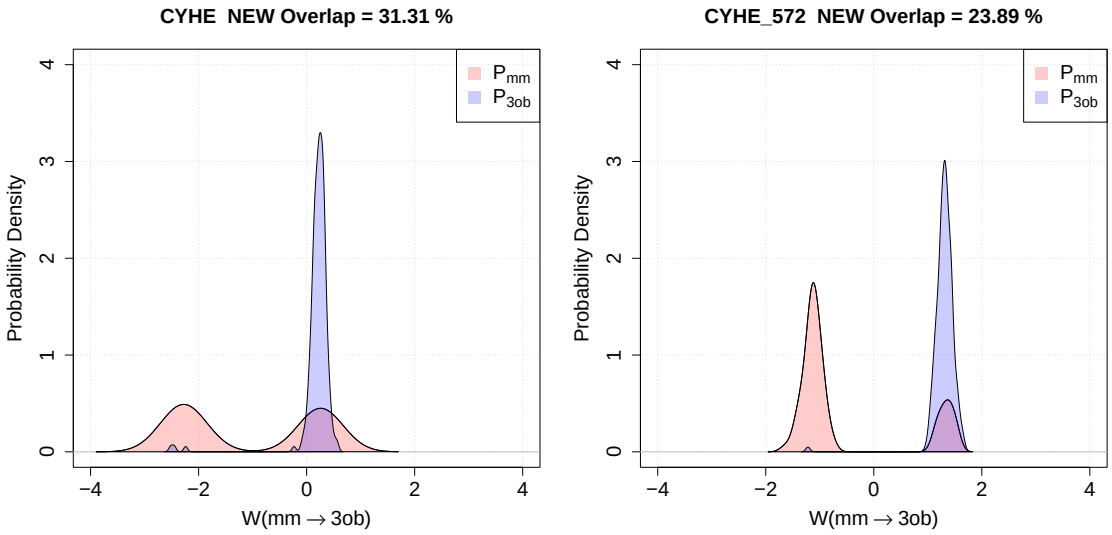
	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
ID	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	$\sigma\Delta A$	W	$\delta\Delta A$
PRPA	-30.97	0.00	0.02	-30.97	0.08	2.45	2.48	30.98	0.00	0.02	30.98	0.09	2.49	2.47	-30.98	0.06	87.18	0.00
cPRPA	-46.65	0.00	0.03	-46.64	0.14	2.38	2.38	46.66	0.00	0.03	46.68	0.14	2.34	2.38	-46.65	0.06	76.28	0.00
BUTA	-82.44	0.00	0.03	-82.43	0.14	2.61	2.38	82.41	0.00	0.03	82.43	0.15	2.61	2.36	-82.43	0.06	84.40	-0.02
cBUTA	-38.96	0.00	0.07	-38.91	0.26	2.14	2.18	38.95	0.00	0.04	39.00	0.26	2.31	2.20	-38.95	0.06	78.85	-0.01
PENT	-33.87	0.00	0.04	-33.84	0.17	2.25	2.31	33.85	0.00	0.04	33.88	0.19	2.48	2.30	-33.86	0.06	78.98	-0.01
cPENT	-59.43	0.00	0.02	-59.42	0.09	2.40	2.46	59.44	0.00	0.01	59.44	0.10	2.42	2.45	-59.43	0.06	82.46	0.01
HEXA	-84.62	0.00	0.04	-84.57	0.23	1.99	2.21	84.55	0.00	0.05	84.59	0.23	2.94	2.24	-84.58	0.06	89.14	-0.04
cHEXA	-7.80	0.31	0.93	-7.35	1.06	2.61	1.38	5.47	0.00	0.03	5.49	0.22	2.61	2.33	-6.23	0.14	23.89	-1.57
cHEXA*	-5.44	0.00	0.03	-5.43	0.12	2.44	2.41	5.44	0.00	0.02	5.45	0.12	2.42	2.41	-5.44	0.06	92.55	0.00

Table 4.9: $\Delta A^{MM\to SQM}_{X,gas}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma\Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations.

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
ID	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	$\sigma\Delta A$	W	$\delta\Delta A$
(1)	-301.75	0.00	0.06	-301.66	0.32	2.11	2.06	301.68	0.00	0.05	301.77	0.33	2.25	2.07	-301.71	0.06	70.82	-0.04
(2)	-869.50	0.01	0.10	-869.37	0.35	1.84	1.96	869.46	0.00	0.06	869.57	0.39	2.25	2.01	-869.46	0.06	68.64	-0.03
(3)	-966.64	0.02	0.14	-966.48	0.37	1.83	1.88	966.55	0.01	0.13	966.69	0.44	2.22	1.92	-966.57	0.07	63.44	-0.06
(4)	-401.66	0.55	0.72	-400.23	0.85	0.48	0.42	400.98	0.06	0.29	402.26	1.42	1.19	0.54	-401.04	0.28	25.05	-0.62
(5)	-438.68	0.00	0.04	-438.65	0.17	2.32	2.33	438.67	0.00	0.03	438.70	0.18	2.38	2.32	-438.67	0.06	81.47	-0.01
(6)	-936.23	0.06	0.28	-935.87	0.52	1.20	1.52	936.03	0.00	0.05	936.44	0.82	2.23	1.44	-936.09	0.07	47.88	-0.14
(7)	-637.96	0.00	0.05	-637.91	0.25	2.14	2.18	637.93	0.00	0.04	637.98	0.26	2.41	2.20	-637.94	0.06	79.40	-0.02
(8)	-740.95	0.00	0.03	-740.93	0.12	2.45	2.41	740.94	0.00	0.02	740.95	0.12	2.46	2.40	-740.94	0.06	86.45	-0.01
(9)	-388.19	0.00	0.02	-388.17	0.16	2.61	2.35	388.15	0.00	0.02	388.16	0.14	2.61	2.37	-388.17	0.06	86.76	-0.02
(10)	-825.15	0.13	0.46	-824.83	0.68	0.40	1.57	825.59	0.10	0.45	826.63	1.91	1.79	0.74	-825.40	0.08	39.67	0.25
(11)	-671.16	0.01	0.09	-671.02	0.39	1.92	1.94	671.13	0.01	0.08	671.26	0.41	2.02	1.94	-671.14	0.07	61.35	-0.02
(12)	-539.12	0.05	0.27	-538.85	0.57	1.76	1.67	539.00	0.05	0.24	539.27	0.58	1.80	1.66	-539.06	0.07	57.29	-0.06
(13)	-98.15	0.37	0.88	-96.38	1.61	-0.57	0.18	99.63	0.54	0.98	102.06	2.32	-0.34	-0.24	-98.93	0.18	9.39	0.77

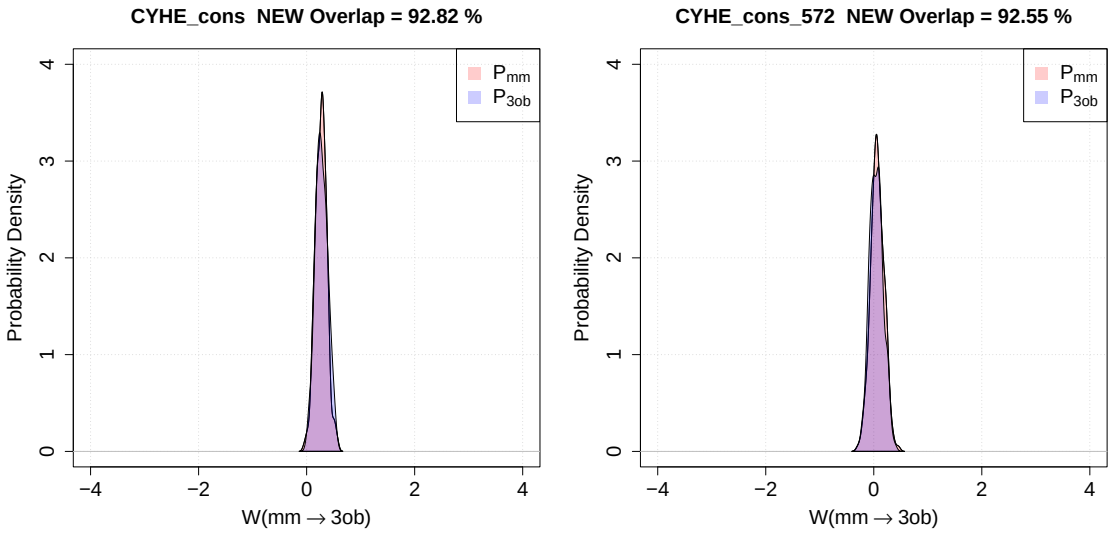
Table 4.10: $\Delta A^{MM\to SQM}_{X,solv}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma\Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations.

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
ID	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	$\sigma\Delta A_i$	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	$\sigma\Delta A$	W	$\delta\Delta A$
(1)	-299.51	0.01	0.10	-299.36	0.41	1.78	1.92	299.39	0.01	0.08	299.62	0.55	2.15	1.74	-299.48	0.07	63.34	-0.03
(2)	-867.22	0.01	0.13	-867.04	0.46	1.62	1.83	867.28	0.01	0.12	867.51	0.56	1.84	1.74	-867.26	0.07	53.39	-0.03
(3)	-964.54	0.04	0.22	-964.25	0.51	1.55	1.64	964.44	0.01	0.11	964.72	0.62	1.93	1.64	-964.47	0.07	51.66	-0.07
(4)	-399.74	0.14	0.45	-398.89	0.90	0.34	0.92	399.77	0.12	0.39	401.39	1.48	0.89	0.28	-399.92	0.11	22.54	0.18
(5)	-446.04	0.02	0.16	-445.25	1.06	0.83	0.99	446.17	0.06	0.31	447.12	1.15	0.86	0.83	-446.18	0.10	28.38	0.13
(6)	-934.40	0.03	0.18	-933.99	0.65	1.33	1.45	934.47	0.01	0.11	934.83	0.72	1.53	1.50	-934.39	0.08	40.76	-0.14
(7)	-634.04	0.09	0.33	-632.80	1.17	0.64	0.57	634.28	0.04	0.26	635.09	1.13	0.67	0.96	-633.94	0.11	22.82	-0.10
(8)	-748.19	0.32	0.57	-746.74	1.24	0.35	0.41	747.93	0.16	0.42	749.66	1.46	0.46	0.21	-748.14	0.12	16.97	-0.04
(9)	-381.27	0.54	0.83	-378.53	1.83	-0.51	-0.42	381.84	0.46	0.69	384.39	1.83	-0.53	-0.31	-381.48	0.20	7.58	0.20
(10)	-824.54	0.49	0.81	-822.95	1.22	0.07	0.31	823.76	0.14	0.47	826.13	2.05	0.68	-0.20	-824.22	0.12	23.29	-0.32
(11)	-673.06	0.03	0.20	-672.46	0.84	1.30	1.19	672.92	0.05	0.27	673.47	0.82	1.32	1.24	-672.96	0.08	40.09	-0.10
(12)	-538.80	0.05	0.25	-538.11	0.83	1.11	1.09	538.75	0.01	0.13	539.28	0.89	1.33	1.28	-538.66	0.09	34.74	-0.14
(13)	-96.91	1.37	1.88	-92.21	3.27	-2.41	-1.36	101.10	1.18	1.46	107.24	3.11	-2.40	-1.93	-99.41	0.43	1.92	2.50



(a) gas phase (b) solvated phase

Figure 4.6: Work distributions of **cHEXA** for switching forward and backward, a) in the gas phase, b) in solution



(a) gas phase (b) solvated phase

Figure 4.7: Work distributions of **cHEXA*** for switching forward and backward when constraining all dihedral angles to remain in the chair configuration, a) in the gas phase, b) in solution

5. Summary & Outlook

Work in the groups of Lee Woodcock and Stefan Boresch established the combination of NEW simulations and JAR as a reliable method for calculating free energy differences between levels of theory. However, the protocols used, e.g., in Ref. [67], are too expensive for many applications. The focus and central result of this thesis is to streamline the existing protocols. It turns out that 200 NEW switches of 2 ps length each (NSWI2000$^{\mathrm{red}}$) are sufficient to achieve converged results in most cases. This amounts to a speedup by a factor of forty compared to the work of [67].

In the gas phase, the optimized NSWI2000$^{\mathrm{red}}$ protocol can be used directly. In aqueous solution, attempting to go directly from the MM to the SQM/MM level of theory may lead to erroneous results. Prolonging the switching length from 2 to 5 ps, a non-negligible increase in computational cost, did not always help. Previous results obtained by Kearns [66] suggested that the source of the poor convergence was the dissimilarity of the solute charge distributions at the two levels of theory. This hypothesis was supported by the work reported in [65]. We, therefore, inserted an intermediate state, described by the regular force field terms, but with averaged Mulliken-like partial charges. Three such approaches were tested; the so-called MULL(solv) hybrid intermediate charge state had the best "price/performance" ratio. In this approach, 200 snapshots stored during the MM simulation of the solvated system, a calculation which is needed anyway, are re-evaluated at the SQM/MM level of theory and the extracted Mulliken charges are averaged. Starting from the MULL(solv) hybrid charge intermediate state, the SQM/MM final state can be reached without convergence problems using the NSWI2000$^{\mathrm{red}}$ protocol. In other words, switches of 2 ps length are sufficient. The savings in computation time compared to, e.g., having to use 5 ps switches more than compensates for the need to also compute $\Delta A^{MM \leftrightarrow MULL}$, the free energy difference between the initial force field description and the hybrid charge intermediate state.

The search for the optimized NSWI2000$^{\mathrm{red}}$ protocol was accompanied by analyses aimed at understanding the sources of slow and/or poor convergence. Using techniques motivated by computational spectroscopy [101], we could show that a minimum switching length of 5 ps is required when trying to go directly from the MM to the SQM/MM level of theory. The same analysis indicated that 2 ps is sufficient when the starting point is a suitable hybrid charge intermediate, such as the MULL(solv) state. These findings justify the conclusions drawn in Ref. [66], which were mostly conjectures at that time. We also developed methods to quickly locate work values in the raw data that deviated significantly from the average, and

in some cases we could associate unusual work values with particular combinations of conformational degrees of freedom. This type of analysis proved extremely helpful in recent work when calculating free energy differences between a system described by MM and its description by a neural net potential [102].

The NEW switches with the NSWI2000$^{\text{red}}$ protocol provide an efficient way to calculate free energy differences between levels of theory, so it is a good point to summarize what obstacles have been overcome and what challenges remain. After Heimdal and Ryde alerted the community that the computation of $\Delta A^{low \to high}$ was far from trivial [37], Essex and co-workers pointed out that small differences in average bond lengths and bond angles when describing a system by, e.g., MM and QM, result in large fluctuations of the energy differences between the two levels of theory. Such broad distributions of energy differences make it practically impossible to obtain converged free energy differences by equilibrium methods. [76] This complication is easily overcome using NEW switching [72, 73, 66, 74, 67, 83, 89]. The NSWI2000$^{\text{red}}$ protocol suggested here should be certainly robust enough to handle any such mismatches, even for relatively large high-level regions.

When using a (S)QM/MM Hamiltonian as the target level of theory, the region of interest is at the initial state treated by MM and by (S)QM at the final state. This may lead to a considerable change of the charge distribution which is seen by the surrounding environment. If, as in the present work, the environment is solvent water, then it takes a finite amount of time for reorientation to occur. The use of intermediate charge states, whether these are averaged Mulliken-derived charges or point charges obtained by some other charge assignment method, provides an efficient way to accelerate such processes. For the TIP3P water model used in this work, we could show that 5 ps are needed if one switches directly from the MM level; this time span can be reduced to 2 ps by using a hybrid intermediate state. These timings may not hold in all cases, but should provide a reasonable guideline whenever there are changes in charge distributions in the region of interest of a multiscale Hamiltonian.

The third cause of slow convergence is different conformational preferences at the two levels of theory. Consider a torsional degree of freedom like the central dihedral angle in butane. Suppose that because of substituents the relative well depths of the trans and gauche conformation are shifted. Specifically, let's assume that at the MM level, the trans state is the minimum energy conformation, but that at the target QM level the gauche conformations are actually lower in energy. A simulation at the MM level of theory will preferentially sample trans conformations; at the QM level of theory, gauche conformations will be more frequent. Ideally, when switching from MM $\to$ QM, a sufficient number of trans conformers will flip

into one of the gauche conformations preferred at the QM end state. If, however, there are high barriers between the trans and gauche minima, very few switching simulations will result in such a conformational change. Thus, in our hypothetical example, most endpoints of MM $\rightarrow$ QM switches will remain stuck in the trans conformation, i.e., the conformation that is typical for the low level, but rare at the high level of theory. Work values obtained during such a series of switching simulations are likely to result in wrong free energy differences. To overcome such conformational preferences may require very long switching lengths, which will be impractical because of computational cost for the foreseeable future.

In this work, the model systems were mostly chosen to avoid this third convergence difficulty; nevertheless, we did encounter some instances of conformational mismatches. For example, for compounds **2** and **7** of the HiPen set, we obtained a few unusual negative work values. As described in I.3.3 of Publication I, there is one dihedral angle in both of these molecules which is exclusively trans at the MM level of theory, whereas at the SQM level of theory the cis conformation is possible. The few switches where a change from trans to cis occurred resulted in the most negative work values. Given the exponential averaging in Jarzynski's equation (Equation 1.5), isolated very negative work values may have a disproportionately large effect on the result. It is quite interesting to note that similar effects for exactly these two compounds were also observed when switching from MM to a neural net potential [102]. In this case, different conformational substates are responsible, but the same "pattern" was present in both NEW switching calculations.

Another example illustrating convergence problems caused by soft degrees of freedom is the unrestrained correction simulations for cyclohexane (cHEXA) reported in 4. First, adequate sampling of all accessible conformations would have required significantly longer equilibrium simulations at the two levels of theory than were carried out in Chapter 4. Second, the prevalence of, e.g., the chair conformation turned out to be distinctly different at the MM and SQM levels of theory. It is not clear what switching lengths would be required for a sufficient number of conformational changes to be observed when switching from MM to SQM; 2 ps switches were clearly far too short. cHEXA, in fact, may prove a challenging test case for any technique that may help overcome convergence problems caused by different conformational preferences at the two levels of theory.

To address such sampling difficulties, two principal approaches may be taken: (1) Ensure adequate sampling both when generating starting configurations for switches, possibly by using enhanced sampling techniques. Obviously, one has to ensure that the snapshots to be used as the starting points for the switches are sampled from

the correct Boltzmann distribution. However, whenever different conformational substates are preferred at the two levels of theory, then this is not sufficient. In this case, one also has to "force" an adequate number of conformational changes during the switches. Relying on such events to "occur by itself" will likely take too much time. At present, it is not clear whether there exist enhanced sampling methods that can be applied during NEW switches and how one can reweigh work values obtained from such biased switches.

The second (2) approach is to also use suitable intermediate states which resemble the target state regarding conformational preferences. The systems we investigated in detail [83, 102] differed at most in two dihedral degrees of freedom. In hindsight, the force field could be easily tweaked to resemble the high level of theory. Such an approach is very similar to the use of intermediate charge states, and was also explored in [65]. The challenge, of course, is what to do for systems with more than two conformational degrees of freedom which may prove problematic and how to locate potential difficulties *before* starting the calculations. One possibility is to systematically modify the force field by an ad-hoc refinement toward the target level of theory. One such approach, force matching, is explored by the group of Lee Woodcock [103, 104, 105].

The results of chapter 4, attempting to improve the computational prediction of solvation free energies by correcting to an SCC-DFTB description of the solute, were at best mixed. In some cases, conformational mismatches as just discussed may have been responsible. In others, e.g., the simple amines **(1)-(4)**, insufficient conformational sampling seems unlikely to be the reason for the poor results. As all convergence criteria we usually monitor gave no indications of convergence problems for these calculations, the only remaining explanation is that SCC-DFTB/MM may not be a good high level of theory, at least for calculating solvation free energies. Given the semi-empirical nature of SCC-DFTB, this may not be too surprising. On the other hand, it has been demonstrated that SCC-DFTB is one of the better SQM approaches regarding reproducing structures and relative energies compared to DFT [38, 39, 40]. Since because of the optimized protocols developed here, correcting free energy differences computed by MM to the SCC-DFTB/MM level of theory is quite efficient, one might add another correction step to a true DFT level of theory. This could be done in one step using Zwanzig's equation or by very short NEW switches. Nevertheless, further research into which (S)QM/MM methods describe, e.g., solute-solvent interactions accurately will be needed.

Abbreviations

AMBER	Assisted Model Building and Energy Refinement
ASFE	Absolute solvation free energies
BAR	Bennett acceptance ratio
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CRO	Crook's Equation
DFT	Density-functional theory
DFTB	Density-functional tight-binding
FEP	Free energy perturbation
FES	Free energy simulation
FF	Force field
GROMOS	GROningen MOlecular Simulation
JAR	Jarzynski's Equation
MAD	Mean absolute deviation
MD	Molecular Dynamics
MM	Molecular Mechanics
NEW	Nonequilibrium work methods
OPLS	Optimized Potentials for Liquid Simulations
PCA	Principal component analysis
QM	Quantum Mechanics
QM/MM	Quantum Mechanical/ Molecular
SCC-DFTB	Self consistent charge density functional tight binding
SQM	Semi-empirical-Quantum Mechanics
SQM/MM	Semi-empirical-Quantum Mechanical/ Molecular Mechanical hybrid methods
TI	Thermodynamic integration

Bibliography

[1] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.

[2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, "Molecular dynamics simulations," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 190–196, 2002.

[3] M. González, "Force fields and molecular dynamics simulations," *École thématique de la Société Française de la Neutronique*, vol. 12, pp. 169–200, 2011.

[4] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1–41, 1995.

[5] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, "CHARMM: the biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.

[6] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53a5 and 53a6," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1656–1676, 2004.

[7] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, no. 6, pp. 1657–1666, 1988.

[8] B. Kirchner, F. Wennmohs, S. Ye, and F. Neese, "Theoretical bioinorganic chemistry: the electronic structure makes a difference," *Current opinion in chemical biology*, vol. 11, no. 2, pp. 134–141, 2007.

[9] S. A. Barr, G. S. Kedziora, A. M. Ecker, J. C. Moller, R. J. Berry, and T. D. Breitzman, "Bond breaking in epoxy systems: A combined QM/MM approach," *The Journal of Chemical Physics*, vol. 144, no. 24, 2016.

[10] K. Vanommeslaeghe and A. MacKerell Jr, "CHARMM additive and polarizable force fields for biophysics and computer-aided drug design," *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1850, no. 5, pp. 861–871, 2015.

[11] M. Nguyen and S. W. Rick, "The influence of polarizability and charge transfer on specific ion effects in the dynamics of aqueous salt solutions," *The Journal of chemical physics*, vol. 148, no. 22, 2018.

[12] M. L. Berkowitz, "Molecular simulations of aqueous electrolytes: role of explicit inclusion of charge transfer into force fields," *The Journal of Physical Chemistry B*, vol. 125, no. 48, pp. 13069–13076, 2021.

[13] L. Hu and U. Ryde, "Comparison of methods to obtain force-field parameters for metal sites," *Journal of Chemical Theory and Computation*, vol. 7, no. 8, pp. 2452–2463, 2011.

[14] J. C. Smith and B. Roux, "Eppur si muove! The 2013 Nobel Prize in Chemistry," *Structure*, vol. 21, no. 12, pp. 2102–2105, 2013.

[15] Q. Cui, T. Pal, and L. Xie, "Biomolecular QM/MM simulations: What are some of the "burning issues"?," *J. Phys. Chem. B*, vol. 125, pp. 689–702, jan 2021.

[16] C. M. Clemente, L. Capece, and M. A. Martí, "Best practices on QM/MM Simulations of Biological Systems," *Journal of Chemical Information and Modeling*, vol. 63, no. 9, pp. 2609–2627, 2023.

[17] N. Hansen and W. F. Van Gunsteren, "Practical aspects of free-energy calculations: a review," *Journal of chemical theory and computation*, vol. 10, no. 7, pp. 2632–2647, 2014.

[18] P. Kollman, "Free energy calculations: applications to chemical and biochemical phenomena," *Chem. Rev.*, vol. 93, no. 7, pp. 2395–2417, 1993.

[19] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, *et al.*, "Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field," *Journal of the American Chemical Society*, vol. 137, no. 7, pp. 2695–2703, 2015.

[20] D. A. Kofke, "Free energy methods in molecular simulation," *Fluid Phase Equilib.*, vol. 228, pp. 41–48, 2005.

[21] R. W. Zwanzig, "High-temperature equation of state by a perturbation method. i. nonpolar gases," *J. Chem. Phys.*, vol. 22, no. 8, pp. 1420–1426, 1954.

[22] W. F. van Gunsteren and H. J. Berendsen, "Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry," *Journal of computer-aided molecular design*, vol. 1, pp. 171–176, 1987.

[23] C. H. Bennett, "Efficient estimation of free energy differences from monte carlo data," *J. Comput. Phys.*, vol. 22, no. 2, pp. 245–268, 1976.

[24] A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free-energy calculations," *J. Phys. Chem. B*, vol. 114, pp. 10235–10253, jul 2010.

[25] A. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, *et al.*, "Best practices for alchemical free energy calculations [article v1. 0]," *Living journal of computational molecular science*, vol. 2, no. 1, 2020.

[26] T. Straatsma and J. McCammon, "Computational alchemy," *Annual Review of Physical Chemistry*, vol. 43, no. 1, pp. 407–435, 1992.

[27] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," *Angewandte Chemie International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.

[28] W. Yang, Q. Cui, D. Min, and H. Li, "QM/MM Alchemical Free Energy Simulations: Challenges and Recent Developments," *Annu. Rep. Comput. Chem.*, vol. 6, pp. 51–62, 2010.

[29] G. König, P. S. Hudson, S. Boresch, and H. L. Woodcock, "Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1406–1419, 2014.

[30] J. Karwounopoulos, Å. Kaupang, M. Wieder, and S. Boresch, "Calculations of absolute solvation free energies with transformato — application to the freesolv database using the cgenff force field," *accepted by JCTC*, 2023.

[31] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles," *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3257–3273, 2012.

[32] A. D. MacKerell Jr, "Empirical force fields for biological macromolecules: overview and issues," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1584–1604, 2004.

[33] A. Hopfinger and R. Pearlstein, "Molecular mechanics force-field parameterization procedures," *Journal of computational chemistry*, vol. 5, no. 5, pp. 486–499, 1984.

[34] B. Lev, B. Roux, and S. Y. Noskov, "Relative free energies for hydration of monovalent ions from QM and QM/MM simulations," *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 4165–4175, 2013.

[35] C. N. Rowley and B. Roux, "The solvation structure of Na^+ and K^+ in liquid water determined from high level ab initio molecular dynamics simulations," *J. Chem. Theory Comput.*, vol. 8, pp. 3526–3535, apr 2012.

[36] Y. Qiu, D. G. Smith, S. Boothroyd, H. Jang, D. F. Hahn, J. Wagner, C. C. Bannan, T. Gokey, V. T. Lim, C. D. Stern, *et al.*, "Development and benchmarking of open force field v1. 0.0—the parsley small-molecule force field," *Journal of chemical theory and computation*, vol. 17, no. 10, pp. 6262–6280, 2021.

[37] J. Heimdal and U. Ryde, "Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations," *Phys. Chem. Chem. Phys.*, vol. 14, no. 36, pp. 12592–12604, 2012.

[38] M. Elstner, K. J. Jalkanen, M. Knapp-Mohammady, T. Frauenheim, and S. Suhai, "Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations," *Chem. Phys.*, vol. 263, no. 2-3, pp. 203–219, 2001.

[39] C. B. Barnett and K. J. Naidoo, "Ring puckering: a metric for evaluating the accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM simulations," *J. Phys. Chem. B*, vol. 114, no. 51, pp. 17142–17154, 2010.

[40] M. Burai Patrascu, E. Malek-Adamian, M. J. Damha, and N. Moitessier, "Accurately modeling the conformational preferences of nucleosides," *J. Am. Chem. Soc.*, vol. 139, no. 39, pp. 13620–13623, 2017.

[41] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Phys. Rev. B*, vol. 58, no. 11, p. 7260, 1998.

[42] M. Elstner, "The SCC-DFTB method and its application to biological systems," *Theor. Chem. Acc.*, vol. 116, no. 1, pp. 316–325, 2006.

[43] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method," *J. Phys. Chem. B*, vol. 105, no. 2, pp. 569–585, 2001.

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

[45] M. Gaus, A. Goez, and M. Elstner, "Parametrization and Benchmark of DFTB3 for Organic Molecules," *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 338–354, 2013. PMID: 26589037.

[46] M. Gaus, X. Lu, M. Elstner, and Q. Cui, "Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1518–1537, 2014. PMID: 24803865.

[47] X. Lu, M. Gaus, M. Elstner, and Q. Cui, "Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 1062–1082, 2015. PMID: 25178644.

[48] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, "Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems," *J. Chem. Theory Comput.*, vol. 11, no. 1, pp. 332–342, 2015.

[49] A. Kumar, P. R. Arantes, A. Saha, G. Palermo, and B. M. Wong, "GPU-enhanced DFTB metadynamics for efficiently predicting free energies of biochemical systems," *Molecules*, vol. 28, no. 3, p. 1277, 2023.

[50] S. I. Allec, Y. Sun, J. Sun, C.-e. A. Chang, and B. M. Wong, "Heterogeneous CPU+ GPU-enabled simulations for DFTB molecular dynamics of large chemical and biological systems," *J. Chem. Theory Comput.*, vol. 15, no. 5, pp. 2807–2815, 2019.

[51] G. Seifert and H. Eschrig, "LCAO-Xα calculations of transition metal clusters," *Phys. Stat. Sol. B*, vol. 127, no. 2, pp. 573–585, 1985.

[52] G. Seifert, H. Eschrig, and W. Bieger, "An approximation variant of LCAO-X-ALPHA methods," *Zeitschrift Fur Physikalische Chemie-Leipzig*, vol. 267, no. 3, pp. 529–539, 1986.

[53] R. S. Mulliken, "Electronic population analysis on LCAO–MO molecular wave functions. I," *J. Chem. Phys.*, vol. 23, no. 10, pp. 1833–1840, 1955.

[54] R. S. Mulliken, "Electronic population analysis on LCAO–MO molecular wave functions. II. overlap populations, bond orders, and covalent bond energies," *J. Chem. Phys.*, vol. 23, no. 10, pp. 1841–1846, 1955.

[55] R. Mulliken, "Electronic population analysis on LCAO–MO molecular wave functions. III. effects of hybridization on overlap and gross AO populations," *J. Chem. Phys.*, vol. 23, no. 12, pp. 2338–2342, 1955.

[56] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations," *Journal of computational chemistry*, vol. 11, no. 6, pp. 700–733, 1990.

[57] L. Cao and U. Ryde, "On the difference between additive and subtractive QM/MM calculations," *Frontiers in chemistry*, vol. 6, p. 89, 2018.

[58] H. Lin and D. G. Truhlar, "QM/MM: what have we learned, where are we, and where do we go from here?," *Theoretical Chemistry Accounts*, vol. 117, pp. 185–199, 2007.

[59] A. O. Dohn, "Multiscale electrostatic embedding simulations for modeling structure and dynamics of molecules in solution: A tutorial review," *International Journal of Quantum Chemistry*, vol. 120, no. 21, p. e26343, 2020.

[60] M. Wieder, M. Fleck, B. Braunsfeld, and S. Boresch, "Alchemical free energy simulations without speed limits. a generic framework to calculate free energy differences independent of the underlying molecular dynamics program," *J. Comput. Chem.*, vol. 43, no. 17, pp. 1151–1160, 2022.

[61] J. Karwounopoulos, M. Wieder, and S. Boresch, "Relative binding free energy calculations with transformato: A molecular dynamics engine-independent tool," *Frontiers in Molecular Biosciences*, vol. 9, sep 2022.

[62] C. Cave-Ayland, C.-K. Skylaris, and J. W. Essex, "A Monte Carlo resampling approach for the calculation of hybrid classical and quantum free energies," *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 415–424, 2017.

[63] A. de Ruiter, S. Boresch, and C. Oostenbrink, "Comparison of thermodynamic integration and bennett acceptance ratio for calculating relative protein-ligand binding free energies," *Journal of computational chemistry*, vol. 34, no. 12, pp. 1024–1034, 2013.

[64] T. J. Giese and D. M. York, "Development of a robust indirect approach for MM→ QM free energy calculations that combines force-matched reference potential and bennett's acceptance ratio methods," *J. Chem. Theory Comput.*, vol. 15, no. 10, pp. 5543–5562, 2019.

[65] S. Ito and Q. Cui, "Multi-level free energy simulation with a staged transformation approach," *J. Chem. Phys.*, vol. 153, p. 044115, jul 2020.

[66] F. L. Kearns, P. S. Hudson, H. L. Woodcock, and S. Boresch, "Computing Converged Free Energy Differences Between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1376–1388, 2017.

[67] F. L. Kearns, L. Warrensford, S. Boresch, and H. L. Woodcock, "The Good, the Bad, and the Ugly: "HiPen", a New Dataset for Validating (S)QM/MM Free Energy Simulations," *Molecules*, vol. 24, no. 4, 2019.

[68] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, no. 14, p. 2690, 1997.

[69] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, no. 3, p. 2361, 2000.

[70] C. Dellago and G. Hummer, "Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics," *Entropy*, vol. 16, no. 1, pp. 41–61, 2014.

[71] H. L. Woodcock, B. T. Miller, M. Hodoscek, A. Okur, J. D. Larkin, J. W. Ponder, and B. R. Brooks, "Mscale: a general utility for multiscale modeling," *J. Chem. Theory Comput.*, vol. 7, no. 4, pp. 1208–1219, 2011.

[72] P. S. Hudson, H. L. Woodcock, and S. Boresch, "Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems," *J. Phys. Chem. Lett.*, vol. 6, no. 23, pp. 4850–4856, 2015.

[73] S. Boresch and H. L. Woodcock, "Convergence of single-step free energy perturbation," *Mol. Phys.*, vol. 115, no. 9-12, pp. 1200–1213, 2017.

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

[75] I. Polyak, T. Benighaus, E. Boulanger, and W. Thiel, "Quantum mechanics/molecular mechanics dual hamiltonian free energy perturbation," *J. Chem. Phys.*, vol. 139, no. 6, 2013.

[76] C. Cave-Ayland, C.-K. Skylaris, and J. W. Essex, "Direct validation of the single step classical to quantum free energy perturbation," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 1017–1025, 2015.

[77] M. A. Olsson, P. Söderhjelm, and U. Ryde, "Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level," *J. Comput. Chem.*, vol. 37, no. 17, pp. 1589–1600, 2016.

[78] M. A. Olsson and U. Ryde, "Comparison of QM/MM Methods to Obtain Ligand-Binding Free Energies," *J. Chem. Theory Comput.*, vol. 13, pp. 2245–2253, Mar. 2017.

[79] C. Steinmann, M. A. Olsson, and U. Ryde, "Relative ligand-binding free energies calculated from multiple short QM/MM MD simulations," *J. Chem. Theory Comput.*, vol. 14, no. 6, pp. 3228–3237, 2018.

[80] K. Vanommeslaeghe and A. D. MacKerell Jr, "Automation of the charmm general force field (cgenff) i: bond perception and atom typing," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3144–3154, 2012.

[81] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell Jr, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3155–3168, 2012.

[82] W. Yu, X. He, K. Vanommeslaeghe, and A. D. MacKerell Jr, "Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations," *J. Comput. Chem.*, vol. 33, no. 31, pp. 2451–2468, 2012.

[83] A. Schöller, F. Kearns, H. L. Woodcock, and S. Boresch, "Optimizing the Calculation of Free Energy Differences in Nonequilibrium Work SQM/MM Switching Simulations," *J. Phys. Chem. B*, vol. 126, no. 15, pp. 2798–2811, 2022.

[84] E. Neria, S. Fischer, and M. Karplus, "Simulation of activation free energies in molecular systems," *J. Chem. Phys.*, vol. 105, p. 1902, 1996.

[85] M. Maroncelli and G. R. Fleming, "Computer simulation of the dynamics of aqueous solvation," *J. Chem. Phys.*, vol. 89, no. 8, pp. 5044–5069, 1988.

[86] M. Maroncelli, "The dynamics of solvation in polar liquids," *J. Mol. Liq.*, vol. 57, pp. 1–37, 1993.

[87] T.-S. Lee, B. K. Allen, T. J. Giese, Z. Guo, P. Li, C. Lin, T. D. McGee Jr, D. A. Pearlman, B. K. Radak, Y. Tao, *et al.*, "Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery," *J. Chem. Inf. Model.*, vol. 60, no. 11, pp. 5595–5623, 2020.

[88] Z. Cournia, C. Chipot, B. Roux, D. M. York, and W. Sherman, "Free energy methods in drug discovery—introduction," in *Free Energy Methods in Drug Discovery: Current State and Future Directions*, pp. 1–38, American Chemical Society, nov 2021.

[89] A. Schöller, H. L. Woodcock, and S. Boresch, "Exploring Routes to Enhance the Calculation of Free Energy Differences via Non-Equilibrium work SQM/MM Switching Simulations Using Hybrid Charge Intermediates between MM and SQM Levels of Theory or Non-Linear Switching Schemes," *Molecules*, vol. 28, no. 10, 2023.

[90] D. L. Mobley and J. P. Guthrie, "FreeSolv: a database of experimental and calculated hydration free energies, with input files," *J Comput -Aided Mol Des*, vol. 28, pp. 711–720, 2014.

[91] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley, "Approaches for calculating solvation free energies

and enthalpies demonstrated with an update of the freesolv database," *J. Chem. Eng. Data*, vol. 62, no. 5, pp. 1559–1569, 2017.

[92] H. Singh and S. Sharma, "Hydration of linear alkanes is governed by the small length-scale hydrophobic effect," *J. Chem. Theory Comput.*, vol. 18, no. 6, pp. 3805–3813, 2022.

[93] L. Pereyaslavets, G. Kamath, O. Butin, A. Illarionov, M. Olevanov, I. Kurnikov, S. Sakipov, I. Leontyev, E. Voronina, T. Gannon, *et al.*, "Accurate determination of solvation free energies of neutral organic compounds from first principles," *Nat. Commun.*, vol. 13, no. 1, p. 414, 2022.

[94] E. Gallicchio, M. Kubo, and R. M. Levy, "Enthalpy- entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation," *J. Phys. Chem. B*, vol. 104, no. 26, pp. 6271–6285, 2000.

[95] S. Jo, T. Kim, V. G. Iyer, and W. Im, "Charmm-gui: a web-based graphical user interface for charmm," *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, 2008.

[96] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, *et al.*, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, 2016.

[97] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLoS computational biology*, vol. 13, no. 7, p. e1005659, 2017.

[98] G. König, F. C. Pickard IV, J. Huang, W. Thiel, A. D. MacKerell Jr, B. R. Brooks, and D. M. York, "A comparison of QM/MM simulations with and without the drude oscillator model based on hydration free energies of simple solutes," *Molecules*, vol. 23, no. 10, p. 2695, 2018.

[99] D. Wu and D. A. Kofke, "Model for small-sample bias of free-energy calculations applied to Gaussian-distributed nonequilibrium work measurements," *J. Chem. Phys.*, vol. 121, no. 18, pp. 8742–8747, 2004.

[100] F. C. Pickard, G. König, A. C. Simmonett, Y. Shao, and B. R. Brooks, "An efficient protocol for obtaining accurate hydration free energies using quantum chemistry and reweighting from molecular dynamics simulations," *Bioorg. Med. Chem.*, vol. 24, pp. 4988–4997, oct 2016.

[101] E. Heid, S. Harringer, and C. Schröder, "The small impact of various partial charge distributions in ground and excited state on the computational stokes shift of 1-methyl-6-oxyquinolinium betaine in diverse water models," *J. Chem. Phys.*, vol. 145, p. 164506, oct 2016.

[102] S. Tkaczyk, J. Karwounopoulos, A. Schöller, H. L. Woodcock, T. Langer, S. Boresch, and M. Wieder, "Using neural network potentials to efficiently calculate indirect free energy estimates," *ChemRxiv*, may 2023.

[103] P. S. Hudson, S. Boresch, D. M. Rogers, and H. L. Woodcock, "Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching," *J. Chem. Theory Comput.*, vol. 14, no. 12, pp. 6327–6335, 2018.

[104] P. S. Hudson, K. Han, H. L. Woodcock, and B. R. Brooks, "Force matching as a stepping stone to QM/MM CB [8] host/guest binding free energies: a SAMPL6 cautionary tale," *J. Comput.-Aided Mol. Des.*, vol. 32, no. 10, pp. 983–999, 2018.

[105] P. S. Hudson, F. Aviat, R. Meana-Pañeda, L. Warrensford, B. C. Pollard, S. Prasad, M. R. Jones, H. L. Woodcock, and B. R. Brooks, "Obtaining QM/MM binding free energies in the SAMPL8 drugs of abuse challenge: indirect approaches," *J. Comput.-Aided Mol. Des.*, pp. 1–15, 2022.

Publications

I. Optimizing the Calculation of Free Energy Differences in Nonequilibrium Work SQM/MM Switching Simulations

Contents

Bibliographic Information

Schöller, A., Kearns, F., Woodcock,H. L., Boresch, S. Optimizing the Calculation of Free Energy Differences in Nonequilibrium Work SQM/MM Switching Simulations *J. Phys. Chem. B* **2022** *126*, 2798–2811. American Chemical Society. DOI: 10.1021/acs.jpcb.2c00696.

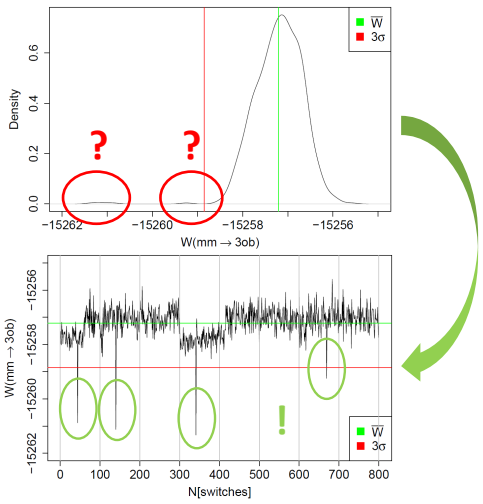
Author's contribution

As first author, I carried out all calculations, made all figures and tables, wrote the first draft of the manuscript and participated in the revisions.

Copyright Notice

Abstract

A key step during indirect alchemical free energy simulations using quantum mechanical/molecular mechanical (QM/MM) hybrid potential energy functions is the calculation of the free energy difference $\Delta A^{low \rightarrow high}$ between the low level (e.g., pure MM) and the high level of theory (QM/MM). A reliable approach uses non-equilibrium work (NEW) switching simulations in combination with Jarzynski's equation; however, it is computationally expensive. In this study we, investigate whether it is more efficient to use more, shorter switches or fewer, but longer switches. We compare results obtained with various protocols to reference free energy differences calculated with Crooks' equation. The central finding is that fewer, longer switches give better converged results. As few as 200 sufficiently long switches lead to $\Delta A^{low \rightarrow high}$ values in good agreement with the reference results. This optimized protocol reduces the computational cost by a factor of 40 compared to earlier work. We also describe two tools / ways of analyzing the raw data to detect sources of poor convergence. Specifically, we find it helpful to analyze the raw data (work values from the NEW switching simulations) in a quasi-time series like manner. Principal component analysis helps to detect cases where one or more conformational degrees of freedom are different at the low and high level of theory.



I.1 Introduction

One of the major challenges in chemistry and biochemistry is elucidating thermodynamics and kinetics of enzyme-substrate interactions at the Ångström level. Force field based methods, in particular molecular dynamics (MD) simulations, have become an essential method in this area. They yield coordinates and velocities of each atom in the system as a function of simulation time and, through statistical mechanics these raw data, can be connected to macroscopic properties at microscopic resolution [1, 2]. One application of particular interest is the calculation of free energy differences (ΔA), the principal determinant whether or not a chemical or biological process will proceed spontaneously [3, 4, 5]. Areas in which free energy simulations are used routinely include the computation of relative binding affinities in drug development [6] as well as the calculation of absolute binding affinities [7]. These methods do not only provide predictions of relative or absolute binding free energy differences, but help to better understand the mechanisms of protein–ligand [8, 9] as well as protein–protein interactions [10, 11, 12]. Whenever a chemical process of interest requires electronic scale details that molecular mechanics (MM) force field cannot provide, the tool of choice is hybrid quantum mechanical/molecular mechanical (QM/MM) calculations [13, 14, 15, 16], for which the Nobel prize was awarded in 2013[17]. The need for QM/MM methods arises in many circumstances, e.g., prediction of pK_a values in complex environments, accurate description of metal ions, and many more [18, 19, 20, 21]. While free energy simulations (FES) with MM force fields has become routine, QM/MM calculations remain computationally demanding, which limits the applicability of QM/MM FES [22, 23]. Furthermore, several technical tricks routinely used in MM calculations of ΔA cease to work at the QM/MM level [24]. Therefore, FES at the QM/MM level of theory at acceptable cost[25] remains an elusive goal.

One common strategy to circumvent these problems is to employ an indirect thermodynamic cycle, which takes advantage of the inherent characteristic of free energy being a state function [24]. Suppose we want to compute a (relative) binding free energy difference $(\Delta)\Delta A_{X\to Y}$ for two ligands X and Y. Instead of computing the required alchemical free energy differences $\Delta A^{high}_{X\to Y}$ directly at a high level of theory, e.g., QM/MM, the alchemical transformation is carried out at a low(er) level of theory, e.g., MM. The cycle is closed by the calculation of the free energy differences between the low and high level of theory at the endpoints. So in order to calculate $\Delta A^{high}_{X\to Y}$, the direct transformation $\Delta A^{low}_{X\to Y}$ and the two correction steps $\Delta A^{low\to high}_{X}$ and $\Delta A^{low\to high}_{Y}$ need to be computed. While calculating the difficult to obtain quantity $\Delta A^{high}_{X\to Y}$ by means of three simpler and cheaper steps seems enticing,

the accurate calculation of the correction steps is crucial and often a cause of failure due to convergence problems and the need for extensive conformational sampling [26, 27]. Further, the need for QM/MM energy/force evaluation during the correction steps makes them per se an expensive task.

Various approaches to compute the corrections $\Delta A^{low \rightarrow high}$ reliably, sometimes referred to as "bookending" corrections, have been suggested [28, 29, 30, 31]. We showed in previous work that the use of non-equilibrium work simulations (NEW)[32] leads to converged values for the correction free energies $\Delta A^{low \rightarrow high}$ where other approaches fail [33, 34, 35]. Since NEW calculations connecting two levels of theory can easily become computationally impractical, it is imperative to keep the cost of such calculations manageable to make real world applications feasible. A related challenge is to find reliable indicators whether a particular NEW switching protocol has resulted in converged results.[27]. In the context of classical mechanical interactions, a large body of work[36, 37, 38, 39] is concerned with optimizing calculations using Jarzinsky's and Crooks'[40] relations. Some studies have attempted to tune switching protocols in NEW simulations [41, 42, 43]. For example, by varying how the coupling parameter, λ, changes during the alchemical FES (rate, as well as shape of the switching protocol) the average work could be minimized, though the effect on statistical error and computational cost of the optimized protocols was less clear [43].

In the present context, i.e., when attempting to compute $\Delta A^{low \rightarrow high}$, the most crucial factor is the computational cost of the force/energy evaluations during the switches since these require calculations at the high level of theory. The overall mitigating factor is that one needs multiple switches, which are completely independent calculations. In other words, the computational cost of NEW switching simulations is a trivially parallelizable task. This brings up the immediate question: is it more efficient to carry out a large number of short switches, or does one achieve better convergence of $\Delta A^{low \rightarrow high}$ with fewer switches of longer simulation length? To formulate a concrete example, which protocol leads to the best converged and most accurate results: 10,000 switches of 100 MD steps each, 1,000 switches of 1,000 MD steps each, or 100 switches of 10,000 MD steps each? On a single processor, the computational cost of the three approaches is identical. If one has a large number of processors/machines available, the first protocol will be computationally the most efficient; however, it may not give the most accurate/precise result. For classical force fields, e.g., Hummer showed early on that fewer, longer switching simulations lead to better convergence and, hence, more accurate results [36]. Recently, Aldeghi et al. carried out an analysis into this question to optimize protocols for the calculation of absolute binding free energies when using classical force fields and suggested switching

lengths of 80 ps.[44] Such switching lengths, however, would be prohibitively costly when employing QM/MM Hamiltonians; ideally, we would want to carry out NEW switches of less than a few thousand MD steps. Since switching lengths over fractions of a picosecond or just a few picoseconds most likely result in endpoints far away from equilibrium, we wanted to ascertain that the findings of, e.g., Hummer[36] hold when computing $\Delta A^{low \to high}$ using — by necessity — very short switching simulations. Therefore, finding a good balance between computational efficiency and accuracy of the results is the primary goal of this study.

In recent work [35] we introduced the "HiPen" test set, a collection of 22 molecules for which we analyzed the convergence of $\Delta A^{low \to high}$. The compounds were selected from the Maybridge Hitfinder library of drug-like compounds [45] based on several criteria. In particular, we picked molecules based on the penalty score when assigning CGenFF 3.0 parameters [46, 47]. Parameter assignment by CGenFF (in particular dihedral angle parameters, partial charges) relies on molecular similarity. A high penalty score indicates that no close model compound could be found in CGenFF's database; therefore, the resulting parameters should be understood as an 'educated guess' only. In the present context, we expected that such "high penalty" force field parameters, at least for some of these compounds, would make it challenging to compute the correction $\Delta A^{low \to high}$, even with NEW based switching methods. Based on the convergence and agreement with reference results, the compounds were classified as "good", "bad," or "ugly". In Ref. [35] a single switching length was used, based on previous experience with model compounds. Similarly, a large number of switches (10,000 per $\Delta A^{low \to high}$) were carried out in an attempt to ensure convergence. Thus, our starting point is to see whether this protocol can be optimized, in particular whether comparable results can be achieved at significantly lower computational cost by varying switching length, number of switching simulations, or both. As in Ref. [35] we only compute $\Delta A^{low \to high}$ for the individual test molecules in the gas phase.

When initially using the simulation protocols described in Ref. [35], we observed a few discrepancies to our earlier results. This prompted us to search for ways to discern and understand factors hindering convergence. In particular, we devised two approaches / tools which not only helped us resolve the deviations from the earlier results, but which are generally useful to spot problems when attempting to compute free energy differences between two levels of theory. First, the raw data needed for Jarzynski's equation [32] are the work values of changing the Hamiltonian from the low to the high level of theory (cf. Theory and Methods). A priori, there is no temporal ordering in these work values. However, the switching simulations

are started from restart files generated from equilibrium simulations at the low of level theory; thus, there is temporal ordering how the starting points were generated. Plotting the work values as if they formed a time series helped us detect that in some cases our equilibration simulations had been too short.

Second, even after repeating the affected calculations with a longer equilibration, for some systems we still observed a (very) small number of switches with outlying work values (i.e., work values deviating significantly from the mean). When we analyzed convergence problems in earlier work [33, 34], we always found strong indications that these were caused from different preferred conformational substates at the two levels of theory. Investigating such effects systematically quickly becomes difficult once multiple dihedral angles are involved. A possibly helpful tool is principal component analysis (PCA) [48], widely used in multivariate data analysis, which not only detects correlations between multiple independent variables, but also displays the relative contribution to the variance for each variable (loadings) in a predictive way. We, therefore, explored whether PCA makes it possible to identify problematic conformational degrees of freedom, e.g., dihedral angles, in a semi-automatic manner.

The remainder of the manuscript is organized as follows. In Theory and Methods we first summarize the theoretical background and introduce the model systems used for validation. Next, we present the technical details of all simulations carried out. Then, we describe the quasi-time series approach to analyze work values from NEW simulations, as well as our use of principal component analysis (PCA). In Results we begin by first showing data obtained based on the protocols from the original HiPen study [35] and analyze cases where convergence was unexpectedly poor using the quasi-time series approach. This led to the use of extended equilibration simulations and the final optimized protocols. We conclude by presenting illustrative results obtained by PCA.

I.2 Theory and methods

I.2.1 Theoretical background

The focus of this study is the free energy difference between two levels of theory, as needed in indirect cycle QM/MM FES. As in our previous work [33, 34, 35], we chose MM as the low level, and the semi empirical (SQM) SCC-DFTB method [49] as implemented in CHARMM[50] as the high level of theory. Our aim is to compute $\Delta A^{MM \rightarrow SQM}$ as accurately as possible while keeping the computational cost low. To obtain these free energy differences, we employed non-equilibrium work (NEW) simulations, using primarily Jarzynski's equation [32]:

$$\Delta A^{MM\rightarrow SQM} = -k_b T \ln \left\langle \exp\left(\frac{-W^{MM\rightarrow SQM}}{k_b T}\right)\right\rangle_{MM} \tag{I.1}$$

In Eq. I.1 k_b is Boltzmann's constant, T the temperature, and W is the work required to change the Hamiltonian from MM to SQM, obtained from, as we call them, "switching simulations", often only referred to as "switches". As required by the theory behind Eq. I.1, switches were started from restart files written at regular intervals during an equilibrium simulation at the MM level of theory at constant temperature and volume, hence the subscript MM. The averaging, indicated by the angular brackets $\langle\rangle$, was carried out over the work values obtained from $N_{replicate}$ switches. This was one of the parameters we varied systematically, the other being N_{switch}, the number of MD steps used per switch (cf. below).

In Ref. [35] results calculated with Eq. I.1 were not always converged. Thus, as in previous work [33, 34, 35], we used Crooks' equation[40] to obtain reference results. The superiority of Crooks' relation to reliably compute free energy differences when there is low overlap between distributions of forward and backward work values, i.e., situations when Jarzinsky's equation is expected to converge poorly, is well documented in the literature.[39, 37, 38, 36] To use Crooks' relation, equilibrium simulations at the high level of theory, as well as switches from the high to the low level of theory ($W^{high\rightarrow low}$) are needed. While this is too expensive for most practical applications, it is relatively affordable when using SQM as the high level of theory, permitting us to get reliable reference values for $\Delta A^{MM\rightarrow SQM}$. Furthermore, several convergence criteria permitting one to estimate the quality of results obtained by Jarzynski's equation require the knowledge of the distribution of forward *and* backward work values (see also A) [27, 35, 51, 37].

I.2.2 Choice of model systems

We already described the HiPen test set in the Introduction. The classification of a molecule as "good", "bad" or "ugly" in Ref. [35] meant that use of Jarzynski's equation was sufficient to obtain converged results for $\Delta A^{MM\rightarrow SQM}$ ("good"), was not sufficient, i.e., only Crooks' equation worked ("bad"), and even use of Crooks' equation did not work ("ugly"). As it seemed futile to attempt optimizing calculations where we encountered convergence problems even when using a very elaborate protocol, in this study we focused primarily on the "good" compounds from Ref. [35]. Therefore, all "good" compounds plus one "bad" compound **21** from the full HiPen set were picked as the test pool for this study.

These compounds are shown in Figure I.1. To ease comparison with Ref. [35]

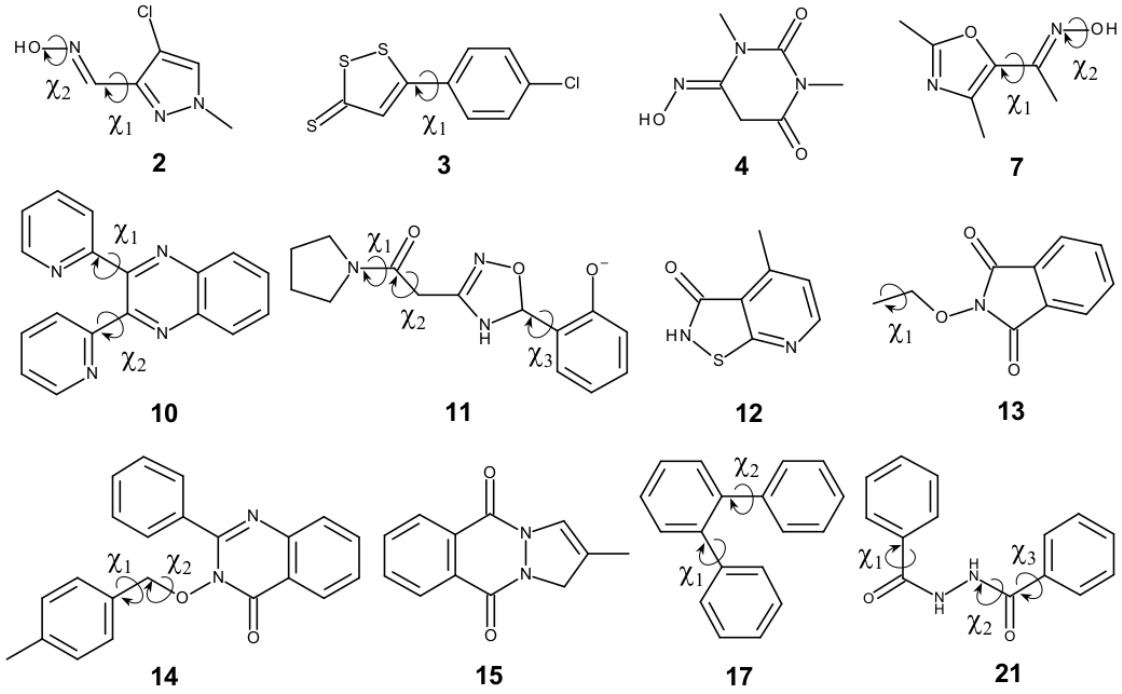


Figure I.1: Subset of HiPen Dataset considered in this work. Dihedral angles, which were candidates for randomization and analysis, are labeled.

we keep the compound IDs used there. The figure also indicates and labels some dihedral angles which we considered to have an influence on the results. Where possible, the dihedral angle labels used in Ref. [35] were kept. Some additional details about the compounds, such as total number of atoms and number of heavy atoms, as well as the respective offset which was subtracted from all $\Delta A^{low \to high}$ results, can be found in Table A.1 of A.

I.2.3 Simulation details

All simulations were carried out with CHARMM (developmental Version 44a2) [52]. All simulations were carried out in the absence of solvent, neither implicit, nor explicit (i.e., gas phase simulations).

Generation of starting points for non-equilibrium switches

Preparation and initial equilibration: Geometry optimized starting coordinates, molecular topologies files and force field parameters were taken from the freely available repository of the HiPen Dataset[1]. For each molecule shown in Figure I.1

[1]DOI:10.5281/zenodo.2328952

eight starting coordinates were generated as follows: Rotatable bonds were randomized, then the coordinates were minimized for 1,000 steps using the adopted basis set Newton Raphson method [53, 54] while restraining the dihedral angles to the respective random values (harmonic dihedral restraints with a force constant of 100 kcal mol^{-1}rad^{-2}). After removing the restraints, Langevin dynamics (LD) was carried out to equilibrate the system at both the MM and SQM levels of theory. A friction coefficient of 5 ps^{-1} was applied to all atoms and random velocities were assigned at each step corresponding to a temperature bath of 300 K. Following the initial protocol from Ref. [35] the length of this equilibration simulation was 10 ps.

For five compounds, **2**, **7**, **10**, **12** and **21**, all simulations described below were repeated using a more elaborate equilibration protocol. First, the length of the equilibration simulation was extended to 5 ns in the MM case and 500 ps in the SQM case. Second, for compound **7**, the randomization of dihedral angles was modified compared to Ref. [35]. In Figure A.1 of A, we show the dihedrals and their labels considered by Kearns et al. and compare them with the present work. As one can see, choosing χ_1^{old} (A, Figure A.1a) simply was an oversight; there is no point in randomizing a dihedral angle in a heterocycle. Similarly, randomizing χ_3^{old} can lead to an isomerization around the C=N double bond (*cis* $\leftrightarrow$ *trans*), which for oximes does not occur under the simulation conditions [55]. Thus, we chose not to randomize these two dihedrals in the modified equilibration, prompting the change in labeling for the dihedrals (cf. Figure A.1 of A). For **7**, this left a single dihedral angle (i.e., χ_2^{old} in the old, χ_1 in the new labeling scheme) as a candidate for randomization. To understand the implications of randomizing χ_1 (χ_2^{old}), we calculated potential energy scans at the MM and 3OB levels of theory (see Figure A.2 of A). The global minimum for dihedral χ_1 is at $\pm 180°$, with a secondary, local minimum at 0°. Given that there are two minima, randomizing χ_1 is an option. However, as the potential energy scans (see Figure A.2 of A) show, the barriers separating the two minima are high, and it seems unlikely that adequate sampling according to their Boltzmann weight would occur during the simulations. For this reason, we decided not to randomize χ_1; it's initial value was set to 180°. Note: the dihedral angle χ_2 indicated in Figure I.1 for compounds **2** and **7** was never considered for randomization; it is included in the figure because it turned out to be relevant for analysis.

MM simulations: 8 LD simulations were carried out, starting from the eight coordinate sets obtained by the preparation and initial equilibration procedure just described. On top of the different coordinates, random initial velocities were assigned in each of the simulations. The simulation length was 10 ns, i.e., 10 million LD steps

with a time step of 1 fs. Restart files were written at every 1,000th step. Thus, during a cumulative simulation length of 80 ns, 80,000 restart files were saved, serving as the pool to carry out non-equilibrium switching simulations to the high (SQM) level of theory. The molecules were fully flexible, and nonbonded interactions were not truncated ("infinite" cutoff radius).

SQM simulations: For each of the eight starting structures equilibrated at the SQM level of theory, a LD simulation of 1 ns (1 million steps with a time step of 1 fs) was carried out. Restart files were written every 100th step, thus resulting in a total of 80,000 restart files generated during a cumulative simulation length of 8 ns. The self-consistent-charge density-functional tight-binding method as implemented in CHARMM [50] with the 3ob-3-1 parameter set [2] [56, 57, 58, 59] was used.

Non-equilibrium work simulations:

Using the restart files written during equilibrium simulations at the MM and SQM levels of theory, non-equilibrium switches MM→SQM and SQM→MM were carried out. As described in detail in earlier work [33], the mixing of Hamiltonians was done with the MSCALE functionality of CHARMM [60] and the work value W during the switch was accumulated with CHARMM's PERT module[52] in slow-growth mode. The time step during the switching simulations was 1 fs. The Hamiltonian was changed linearly from MM to SQM ("forward") and SQM to MM ("backward") during a period of 200 fs (200 steps, NSWI200), 500 fs (500 steps, NSWI500), 1000 fs (1,000 steps, NSWI1000) and 2000 fs (2,000 steps, NSWI2000); cf. Table I.1. A "reduced" protocol, in which we used only a quarter of the switches generated for NSWI2000, is referred to as NSWI2000$^{\mathrm{red}}$.

Depending on the switching length, N_{switch}, the non-equilibrium simulations were launched only from every 10^{th} (NSWI200), 25^{th} (NSWI500), 50^{th} (NSWI1000) and 100^{th} (NSWI2000) restart file saved during the equilibrium simulations. This results in the number of switches summarized in Table I.1. Switches were carried out in both the forward and backward direction. While the focus of this work is on the forward (MM to SQM) direction using Jarzynski's equation (Eq. I.1), the backward work values were needed for histograms and densities of work distributions in the two directions, as well as to compute reference values using Crooks' equation. The densities, as shown, e.g., in Figure I.2, were calculated with R (version 3.4.4) [61] using the built-in `density()` function based on kernel density estimation, without adjusting the bandwidth manually or giving any additional parameters (density.default) [62,

[2] https://www.dftb.org/parameters/download/3ob/3ob-3-1-cc/

Table I.1: Combinations of switching length N_{switch} in fs and number of switches ($N_{replicate}$) studied. Except for NSWI2000^{red}, in which a *red*uced subset of the NSWI2000 data was used, the cumulative length of all switching steps t_{total} is identical in all cases to keep computational cost constant. The column 'Blocksize' indicates the number of switches started from restart files saved during each of the eight equilibrium simulations carried out for each system.

	N_{switch} [fs]	$N_{replicate}$	Blocksize	t_{total} [ns]
NSWI200	200	8000	1000	1.6
NSWI500	500	3200	400	1.6
NSWI1000	1000	1600	200	1.6
NSWI2000	2000	800	100	1.6
NSWI2000^{red}	2000	200	25	0.4

63, 64, 65, 66].

The major goal of this study is to search for an optimal combination of switching length N_{switch} and number of switches $N_{replicate}$, or, phrased differently, to answer the question whether it is better to use many short switches or fewer longer switches. The combinations of N_{switch} and $N_{replicate}$ listed in Table I.1 make possible such a comparison as the cumulative number of simulation steps is identical for each of the four cases; i.e., on a single processor the computational effort would be identical. In the following, the shorthands NSWI200, NSWI500, etc. will not only be used to indicate switching lengths of 200 fs, 500 fs, etc., but to denote the respective protocols (combination of switching length and number of switches) shown in Table I.1.

I.2.4 "Quasi time series" analysis

The distribution of work values obtained from forward and backward switches is often far from Gaussian; quite frequently it is even multimodal. A prototypical example is shown in the top plot of Figure I.2; the data are for molecule **2** using the 10 ps equilibration simulation. The red circles indicate two clusters of work values which are noticeably more negative from the mean $\bar{W}$ (green line). The red line indicates $\bar{W} - 3\sigma$; consequently, more negative work values are expected to occur rarely. Since highly negative work values have a significant weight in Jarzynski's equation Eq. I.1, they might be responsible for systematic deviations between results obtained by Jarzynski's and Crooks' equation, respectively, or be the reason for slow convergence.

From a distribution of work values, details about outliers leading to additional modes and/or shoulders in the main mode are difficult to discern. Ideally, the work

values one uses are statistically independent. In the present work, the NSWI200 switches were started from restart files written in 10 ps intervals, for the NSWI2000 switches the interval was 100 ps (cf. above). When computing the average in Jarzynski's equation the ordering of the work values is irrelevant. Nevertheless, the generation of the restart files introduces a partial temporal ordering. The eight simulations during which the restart files were written are indeed statistically independent (different initial velocities *and* coordinates). Within each of the simulations, however, restart files were written out in order and statistical independence cannot be guaranteed. We, therefore, decided to treat and plot our data (work values) as if they formed a time series.

This is shown in the bottom plot of Figure I.2. The eight "sections" separated by thin gray lines correspond to the eight independent simulations during which restart files were written, we refer to these as "blocks". Within such a section/block the work values are plotted in the temporal order in which the restart files were saved. The example shown in the figure is for the NSWI2000 data, i.e., 100 switches were started from the full set of restart files saved during the respective equilibrium simulation. Rather than resetting the counter between independent simulations, we number the switches from 1–100 for the first simulation, 101–200 for the second simulation etc., as shown on the x-axis of the plot. We refer to this representation of the work values as a "quasi time series". As in the density plot (top of Figure I.2), $\bar{W}$ is indicated as a green and $\bar{W} - 3\sigma$ as a red line. Plotting the data in such a manner automatically leads to a simple labelling scheme for the individual switches; e.g., switch 341 is the 41th NSWI2000 switch started from restart files saved during the fourth out of the eight independent equilibration simulations.

A plot as shown at the bottom of Figure I.2 now makes it straightforward to pinpoint the switches leading to the effects reflected in the corresponding density plot. In particular, one can easily identify four work values (highlighted by green circles), which deviate from $\bar{W}$ by more than 3σ. In the following, we will often refer to work values with $W < \bar{W} - 3\sigma$ as 'outliers'. The threshold $\bar{W} - 3\sigma$ is somewhat arbitrary, but it is an easily applied criterion to automatically detect and flag highly negative work values. The occurrence of such 'outliers' does not indicate that something is wrong; given a sufficiently large number of NEW switches, such negative work values are to be expected. Nevertheless, it may be insightful to understand why a particular switch (switching path) results in a significantly more negative work value. In the quasi time series in Figure I.2 one further sees two regions with work values somewhat more negative than $\bar{W}$; these are responsible for the slight shoulder towards more negative work values in the main peak of the distribution of work

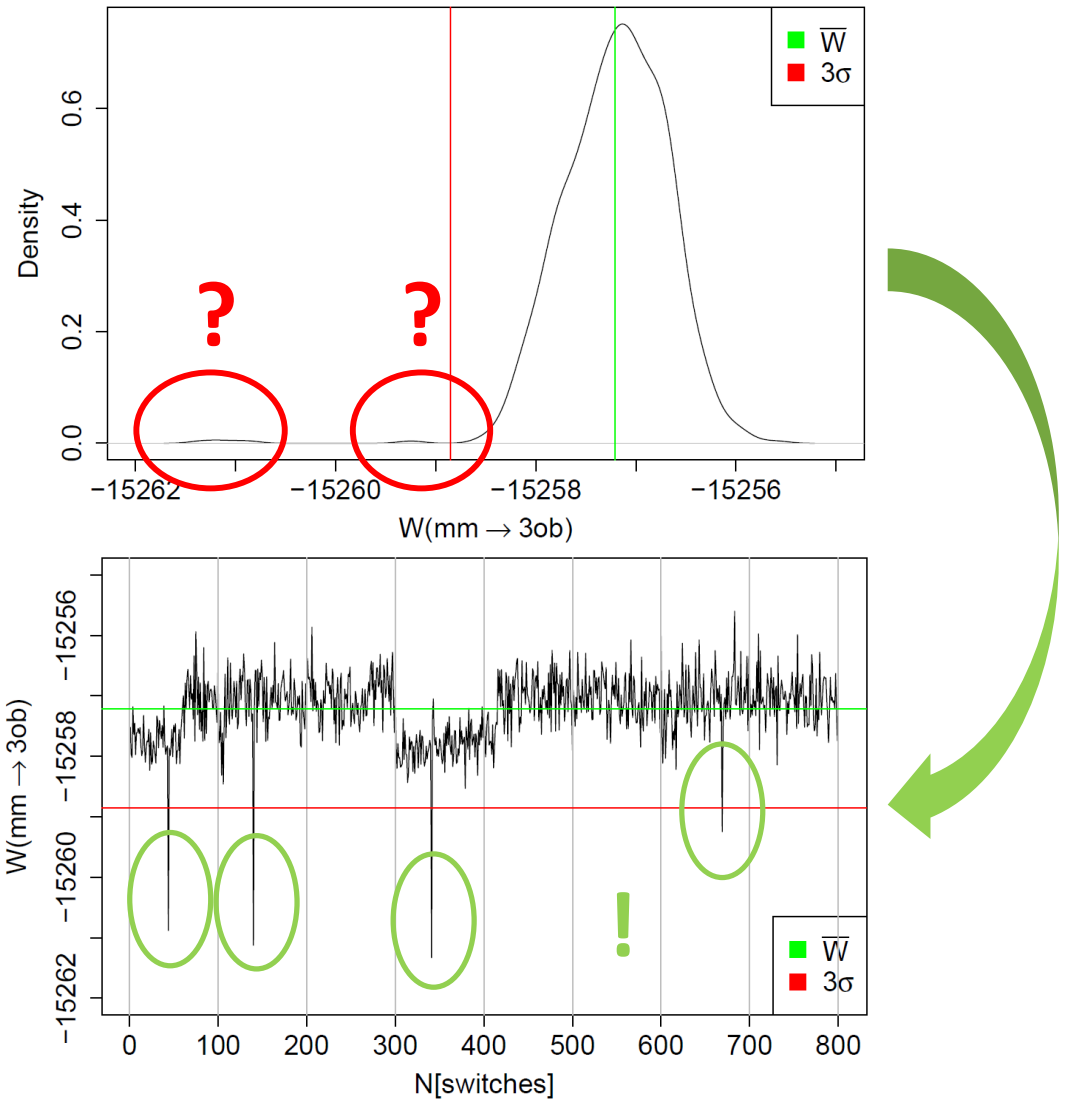


Figure I.2: Top: example of a non-Gaussian distribution of $MM \rightarrow SQM$ work values W. Additional modes of more negative work values are circled in red. The average work value $\bar{W}$ is indicated by a green line, $\bar{W} - 3\sigma$ by a red line. Bottom: the same data displayed as a "quasi" time series, green and red lines indicate $\bar{W}$ and $\bar{W} - 3\sigma$ as before. Individual switches with work values deviating by more than 3σ are highlighted by green circles. All work values are in kcal/mol.

values.

I.2.5 Principal component analysis (PCA):

To investigate the factors resulting in outliers, we used principal component analysis (PCA). As raw data we used the work values and selected values of dihedral angles (cf. Figure I.1) before (χ_i^{pre}) and after the switch (χ_i^{post}). PCA was carried out with

R (version 3.4.4) [61] using the built-in `pcrcomp()` function, which calculates the principal components via singular value decomposition, either on the unscaled or the scaled data matrix [62, 67, 66]. Scaled PCA was employed for collective variables with different value-regimes (e.g., work values W and dihedral angles χ_i); the unscaled version was used when including dihedral angles only. To gauge consistency, the cumulative sum of the principal components' variance quantity was always calculated. Generating the input for the PCA is trivial and the computational effort is negligible. The work values are available anyways, and the CHARMM input scripts for carrying out the switching simulations were modified to save selected dihedrals before and after the non-equilibrium molecular dynamics run.

I.3 Results

I.3.1 Optimization of switching simulations

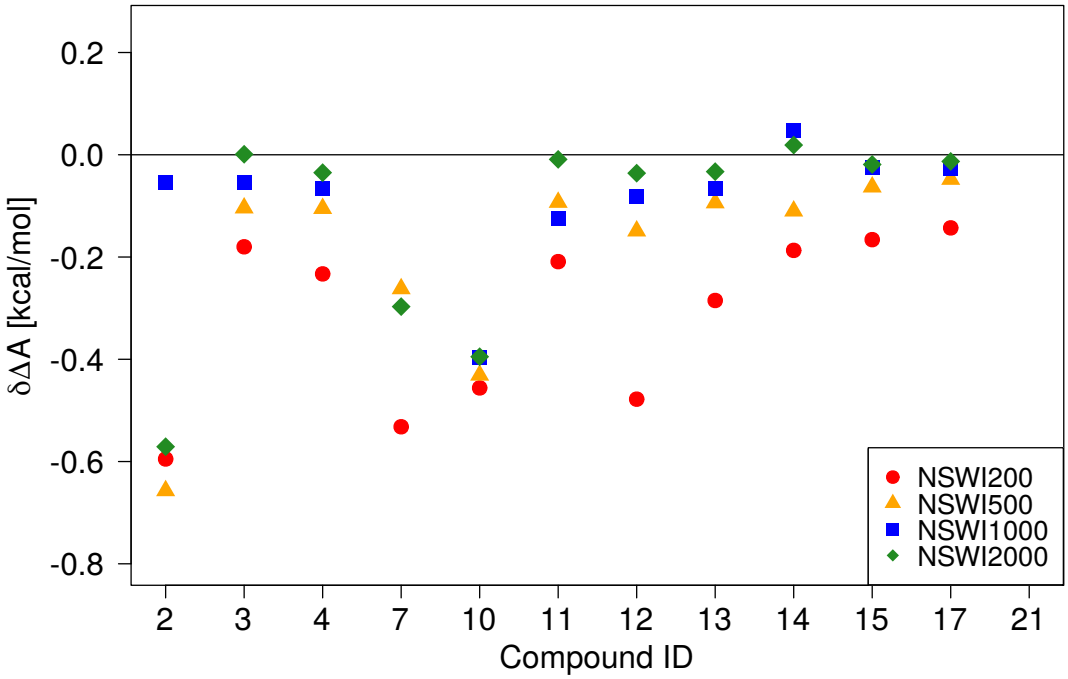


Figure I.3: $\delta\Delta A$ (Eq. I.2) for the results of the forward switching simulations (MM $\rightarrow$ SQM) using the original equilibration protocol as a function of switching length. Some values are off-scale; Figure A.3 in A shows the same data including all values.

Figure I.3 shows the results of the NEW-switching simulations in the forward (MM $\rightarrow$ SQM) direction obtained with the original short equilibration protocol. Each

data point represents the difference

$$\delta\Delta A^{MM\rightarrow SQM} = \Delta A_{JAR}^{MM\rightarrow SQM} - \Delta A_{CRO}^{MM\rightleftarrows SQM} \tag{I.2}$$

between the free energy difference obtained by Jarzynski's equation $\Delta A_{JAR}^{MM\rightarrow SQM}$ for a particular switching length and the best reference result obtained by Crooks' equation $\Delta A_{CRO}^{MM\rightleftarrows SQM}$. Figure A.4 in A is identical to Figure I.3 but also includes error estimates. For detailed results of all conducted switching lengths, including convergence metrics, see supplementary material Tables A.3–A.6. These tables also list the results obtained by Crooks' relation for the switching protocols of different length; as one can see, there is almost no variation with switching length, indicating that these results are well converged. Some results are not included in Figure I.3 as they are off-scale (specifically the NSWI1000 result for **7** ($\delta\Delta A = -1.6$ kcal/mol, and all results for **21** ($\delta\Delta A =$ -6.1 kcal/mol, -4.6 kcal/mol, -3.3 kcal/mol and -1.0 kcal/mol for NSWI200, NSWI500, NSWI1000 and NSWI2000, respectively); these will be discussed shortly.

One can clearly see that in most cases the longest switching protocol (NSWI2000, green diamond) is closest to the respective reference result. Illustrative examples are, e.g., **3** and **4**, where in terms of $\delta\Delta A$, NSWI2000 < NSWI1000 < NSWI500 < NSWI200; i.e., using 800 switches of 2000 fs leads to better results than 1600 switches of 1000 fs etc. The ordering is not always perfect; e.g., for **11** the NSWI500 result is slightly better than NSWI1000; nevertheless, the lowest-magnitude $\delta\Delta A$ is obtained with the NSWI2000 protocol. While there are exceptions (**2**, **7**, **10** and **21**), which will be analyzed next, Figure I.3 strongly suggests that it is more efficient to conduct fewer longer switches than many short ones. In most cases, 800 2ps switches delivered a more reliable result than 8,000 200fs switches while requiring the same computational effort.

To better understand the problematic cases, we present more detailed plots of $\delta\Delta A$ as a function of switching length. In FigureI.4 we plot $\delta\Delta A$ vs. N_{switch} for compound **12**, our reference which has ideal behavior, as well as for three of the four outliers **2**, **10** and **21**. These represent different "types" of deviation. The fourth outlier, **7**, behaved somewhat similarly to **2**. Both **2** (Figure I.4b) and **7** do not obey the general trend that increasing switching length reduces $\delta\Delta A$. For **2** NSWI1000 gives an almost perfect result whereas the longer NSWI2000 protocol deviates from the reference result by almost as much as NSWI200 and NSWI500. In the case of **7** NSWI500 and NSWI2000 perform well, but NSWI1000 deviates significantly from the reference result. Compound **10** exhibits the curious behavior of a systematic offset compared to the reference result, which does not become noticeably smaller

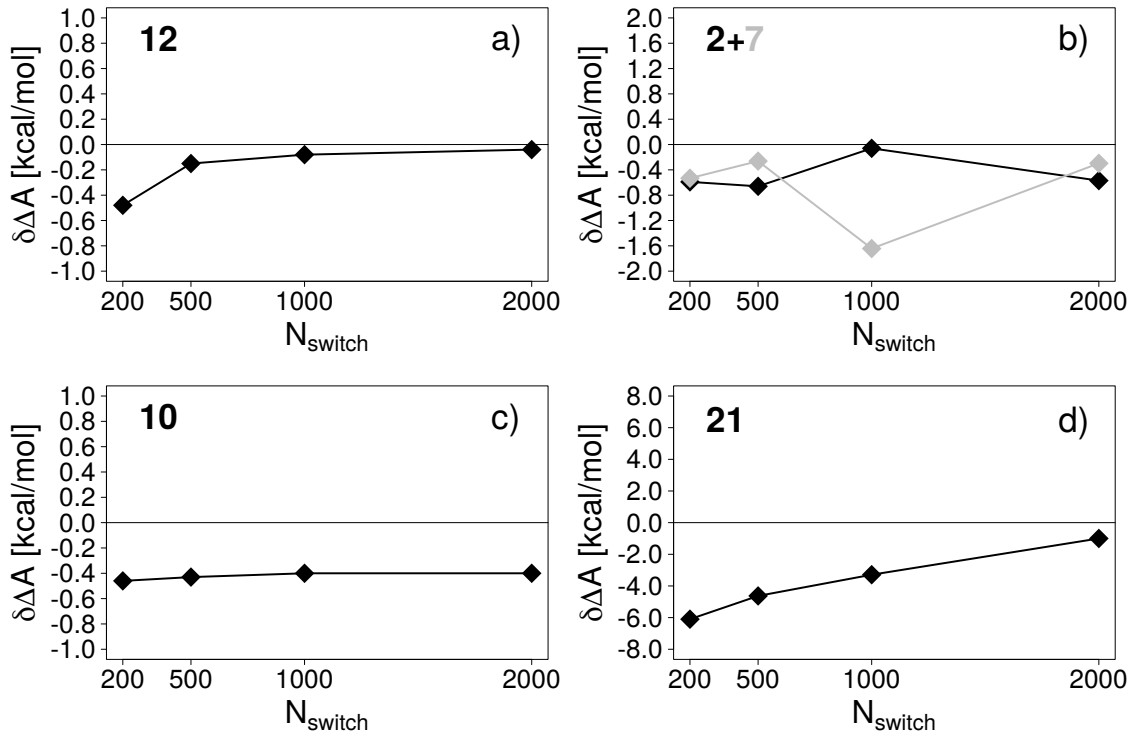


Figure I.4: Detailed dependence of $\delta\Delta A$ on switching length for a) compounds **12** b) **2** and **7**, c)**10** and d) **21**.

as the switching length increases and remains at 0.4 kcal/mol even for NSWI2000 (Figure I.4c). Finally, at first glance **21** behaves as expected (Figure I.4d); i.e., $\delta\Delta A$ becomes smaller as the switching length increases, but even when using switching lengths of 2 ps, the deviation to the reference result remains unacceptably high (≈ -1 kcal/mol). In principle, the result for **21** was not surprising as it was classified as a *bad* compound.

I.3.2 "Quasi time series" analysis of switching simulations

Our starting point to understand why the results for compounds **2**, **7**, **10** and **21** failed to improve when increasing N_{switch} was to scrutinize their forward and backward work distributions. For all of them, the distributions were far from Gaussian, and in all cases there was a non-negligible number of switches with distinctively more negative work values. An illustrative example is shown in Figure I.2. Since these outliers have high weight in Jarzynski's equation, these were likely responsible for the systematic deviations of the forward results from the reference values obtained using Crooks' equation.

An obvious question to ask is whether these outliers occur at random, or whether

there is some correlation to the order in which the starting coordinates (restart files) for the switches were generated. Therefore, as described in the subsection "Quasi time series" Analysis, we plotted the work values as a quasi time series. Illustrative examples for the initial simulation protocols ($N_{switch} = 2000$) are shown in Figure I.5a for **10** and Figure I.6a for **21**. The eight blocks of data (restart files) are marked by the vertical gray lines.

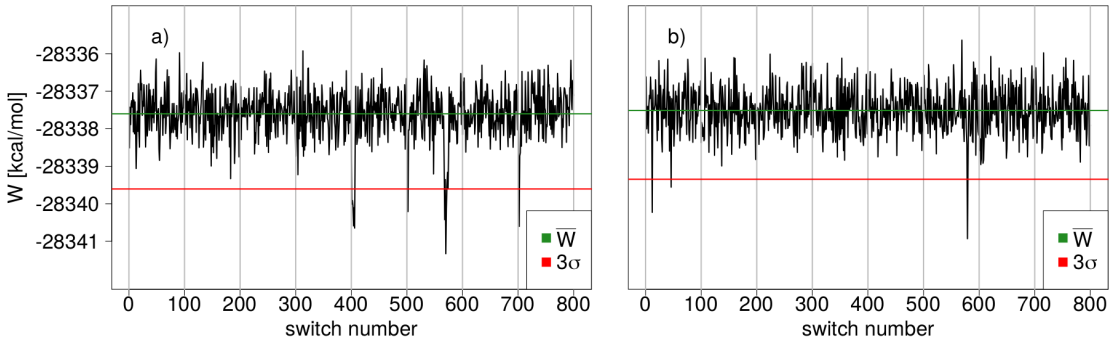


Figure I.5: Quasi time series of Compound **10** plotted as W [kcal/mol] for the 2000 fs switches vs. switch index, concatenating the underlying equilibrium simulations for a) the original protocol and b) for the modified protocol.

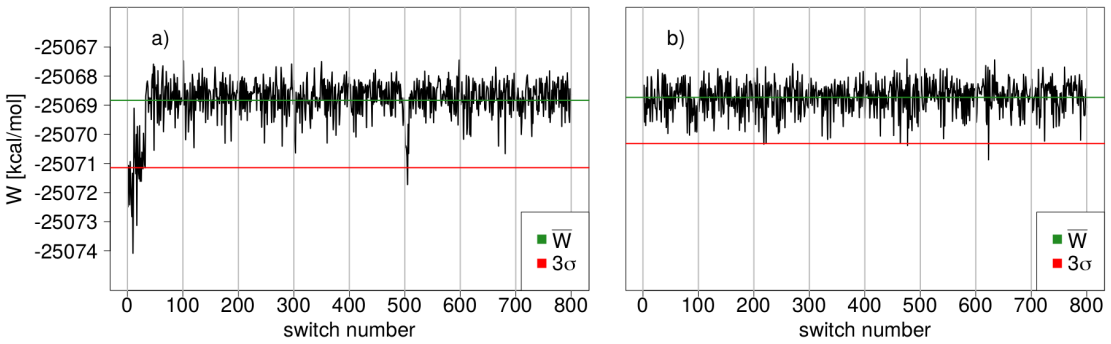


Figure I.6: Quasi time series of Compound **21** plotted as W [kcal/mol] for the 2000 fs switches vs. switch index, concatenating the underlying equilibrium simulations for a) the original protocol and for b) the modified protocol.

Focusing on negative outliers, work values $W_i < \overline{W} - 3\sigma_W$ (see the -3σ line in red in Figures I.5 and I.6), one notices that they occur often at the beginning of a replica, i.e., at the beginning of one of the eight independent sets of starting points. E.g., for **10** there are outliers near switch number 400, 500 and 700 (see Figure I.5a), and for **21** there is a sizeable number of switches with significantly more negative work values starting at switching number 0, as well as few values near switching

number 500 (see Figure I.6a). In other words, several switches started from restart files generated early during the respective equilibrium simulation at the MM level of theory led to unusually negative work values. Note that this is not always the case, e.g., for **10** several switches after switch number 560 also give low work values (see Figure I.5a). This observation prompted us to scrutinize our protocol for system preparation and equilibration (cf. Generation of starting points for non-equilibrium switches), and we repeated the full sequence of simulations for **2**, **7**, **10** and **21** with the (much) longer equilibration simulations following the random dihedral angle assignment. To ensure that this did not affect the unproblematic cases, the longer protocol was also applied to **12**.

Table I.2: MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) of results obtained from forward switches and Jarzynski's equation compared to Reference result (CRO) including corrected results for **2**, **7**, **10** and **21**.

	MAD [kcal/mol]	RMSE [kcal/mol]
NSWI200	0.27	0.31
NSWI500	0.08	0.09
NSWI1000	0.08	0.09
NSWI2000	0.04	0.05
$\text{NSWI2000}^{\text{red}}$	0.06	0.08

The effect on the quasi time series of the work values is displayed for **10** and **21** in Figs. I.5b and I.6b, respectively. For **21** all negative outliers have effectively disappeared, for **10** their occurrence is much rarer. While we do not show the plots for **2** and **7** (see Figs. A.6, A.7 in A), the overall picture is very similar. Occasional negative outliers, work values with $W_i < \overline{W} - 3\sigma_W$, remain, but their frequency is reduced. The modified protocol has a dramatic quality of the overall results, as can be seen in Figure I.7 and Table I.2. Now, the longest switching protocol NSWI2000 leads to the lowest deviation $\delta\Delta A$ from the reference in almost all cases; any deviations, such as for **7**, where $\delta\Delta A(\text{NSWI1000}) < \delta\Delta A(\text{NSWI2000})$, are so small as to be irrelevant in practice. Most surprisingly, **21** for which we had obtained large $\delta\Delta A$ values for all switching lengths, in line with its initial classification as "bad", is now "perfectly behaved". Full results for all compounds repeated with the longer equilibration protocol can be found in Tables A.3–A.6 of A, and Figure A.5 in A shows the data in Figure I.7 with error bars. Figure A.8 of A displays the convergence behavior as a function of switching length, analogously to Figure I.4 above, for **12**, **2**, **10** and **21**.

Having identified the NSWI2000 protocol as the most reliable method to compute

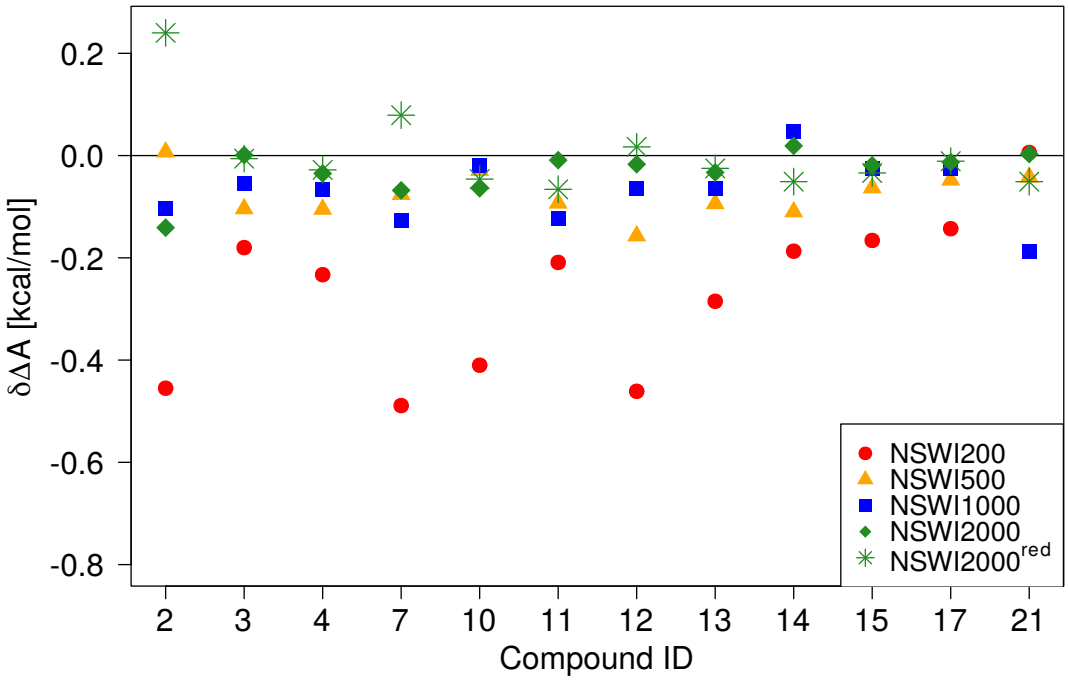


Figure I.7: $\delta\Delta A$ calculated as the difference between two sided method CRO and one sided method JAR for switching forward($MM \rightarrow SQM$) after correcting the equilibration period for **2**, **7**, **10**, **12** and **21** from 10 ps to 5 ns for MM and from 10 ps to 0.5 ns for SQM after random dihedral assignment procedure.

free energy differences $\Delta A^{low \rightarrow high}$ using Jarzynski's equation, we attempted to reduce the computational effort further. Specifically, we lowered the number of switches per block from 100 to 25, i.e., using $N_{replicate} = 200$ instead of 800. These results are labeled NSWI2000$^{\text{red}}$ and are included in Figure I.7 (green stars) and Table I.2. While the MAD of NSWI2000$^{\text{red}}$ is slightly larger than for NSWI2000, it is still smaller than MAD(NSWI1000) at one quarter of the computational cost. As can be seen in Figure I.7, most $\delta\Delta A$ values remain almost unchanged though, e.g., for **2** there is already some deterioration.

The use of quasi time series analysis suggested corrections to our simulation protocol and dramatically improved the convergence of the computed $\Delta A^{MM \rightarrow SQM}$ values. Nevertheless, as one sees, e.g., in Figure I.5b, even when employing the longer equilibration protocol some work values remain outliers, based on our criterion of deviations by more than three standard deviations from $\bar{W}$. In principle, considerably more negative work values could bias the free energy difference computed by Jarzynski's equation. To check whether this is the case here, we recomputed $\Delta A^{low \rightarrow high}$ using Jarzynski's equation for **2**, **7**, **10**, and **21**. The results for the work

values from the NSWI2000 and NSWI2000$^{\text{red}}$ switching protocols with the outliers present or removed are shown in Table I.3.

Table I.3: Outlier Impact on JAR(fw) for **2**, **7**, **10** and **21** for NSWI2000 and NSWI2000$^{\text{red}}$ with and without outliers. The results are displayed as $\delta\Delta A$ in kcal/mol, similarly to Figure I.3 and Figure I.7. $N_{Outliers}$ is the number of outliers identified in the time series like analysis in the NSWI2000 and NSWI2000$^{\text{red}}$ raw data. A * denotes results obtained with all outliers excluded from the calculation.

	$N_{Outliers}$		$\delta\Delta A$			
	NSWI2000	NSWI2000$^{\text{red}}$	NSWI2000	NSWI2000*	NSWI2000$^{\text{red}}$	NSWI2000$^{\text{red}}$*
2	4	0	-0.14	0.26	0.24	0.24
7	4	1	-0.07	0.10	0.08	0.10
10	4	1	-0.06	0.12	-0.06	0.09
21	5	1	0.01	0.04	-0.05	0.01

For **7**, **10**, and **21**, the difference in $\delta\Delta A$ with and without the outliers is less than 0.2 kcal/mol, which is negligible in practice. The free energy difference is shifted slightly to more positive values, as expected. For each of these three compounds, the reduced set of work values NSWI2000$^{\text{red}}$ still contains one of the outliers. For **2**, the difference between the NSWI2000 (outliers included) and NSWI2000* (outliers excluded) results is 0.4 kcal/mol. While not negligible anymore, this is most likely still acceptable. More importantly, however, for **2** the reduced NSWI2000$^{\text{red}}$ set does not contain a single outlier. As seen in Figure I.7, **2** is the single case where using just 200 work values affected the result noticeably; this is the result of the reduced data set incidentally not containing any of the 'outliers'.

I.3.3 PCA of switching simulations

In earlier work [34, 35] we showed that different conformational preferences at the two levels of theory had a noticeable impact on convergence. It, therefore, seemed likely that the outliers in terms of more negative work values were somehow correlated with dihedral degrees of freedom. We used PCA to investigate this question systematically and to identify the specific degree(s) of freedom responsible. As described in Principal Component Analysis, our input variables are the dihedral angles considered relevant before and after the switching simulations (see Figure I.1), as well as the work values themselves. In the following we focus on compounds **2** and **7**. For both systems, even using the extended equilibration protocol, we obtained isolated outliers in the work values (deviating by more than 3σ from $\bar{W}$).

Results of the PCA analysis for **2** are summarized in Figure I.8. The data shown

!ht]

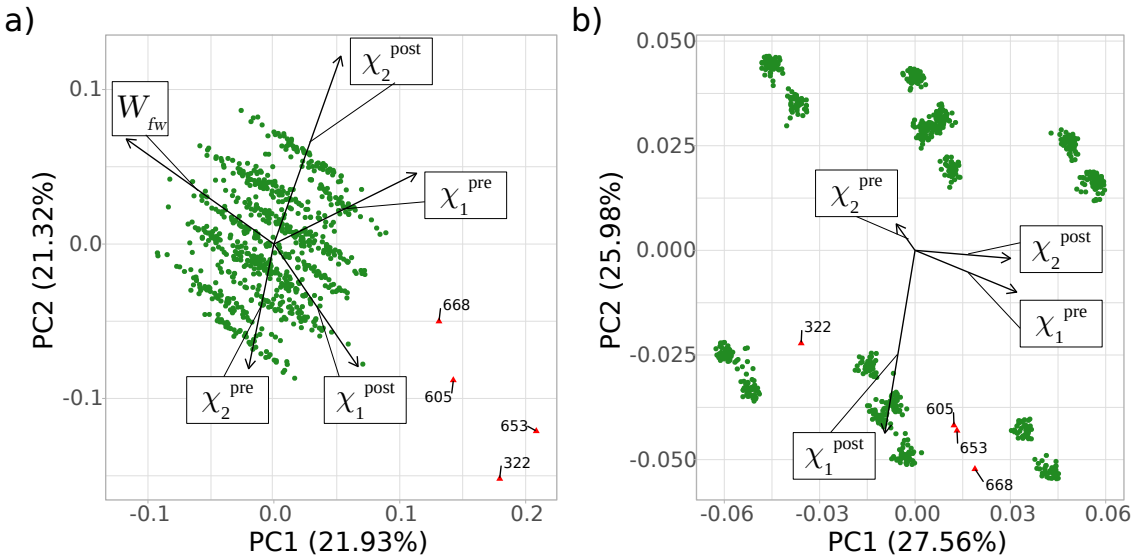


Figure I.8: PCA of compound **2** plotted as PC1 vs. PC2 for the longer equilibration protocol as a) scaled version and b) unscaled version. Outliers are labelled explicitly and marked by red triangles (rather than green dots) according to the predefined outlier criteria (deviations less than $\bar{W} - 3\sigma$ for a), and $|\bar{\chi}_2^{post}| - 3\sigma$ for b), cf. Quasi time series analysis).

are for the final, long equilibration protocol; the plots for the original protocol can be found in A (Figure A.9). The corresponding quasi time series analysis is shown in Figure A.6 of A. Figure I.8a is the scaled PCA, in which dihedral values χ_1, χ_2 and work W were used as inputs. For the dihedral angles we distinguish the values before (denoted as χ^{pre}) and after (χ^{post}) the switch. Highlighting of outliers, colored in red rather than green, and with explicit labels added, was done in an automated manner during the calculation/plotting of the PCA results according to the predefined criteria with respect to deviation from $\bar{W}$ or $|\bar{\chi}_2^{post}|$; cf. Quasi Time Series Analysis. The arrows in the plot are the loading vectors describing which feature of the input data correlates with the calculated principal component and in which way (positive/negative).

We start with the scaled PCA plot for **2** shown in Figure I.8a. Four switches (322, 605, 653, and 668) behave very differently; based on outlier criterion $\bar{W} - 3\sigma$ they were colored in red automatically. These are exactly the switches which can be easily discerned in the quasi time series analysis; see Figure A.9b) in A. The four switches are located in the opposite direction of the loading vector for W indicating an unusual negative deviation of the feature W. Besides W, χ_2^{post} has the largest loadings vector among all the variables, suggesting that this conformational degree

of freedom may be of relevance as it accounts for a large proportion of the variance. In the unscaled PCA, shown in Figure I.8b, we used $|\bar{\chi}_2^{post}| - 3\sigma$ as the criterion for outlier detection (cf. Quasi Time Series Analysis). The only switches highlighted by the automated coloring are 322, 605, 653, and 668. This demonstrates that the four negative values in W are correlated and, possibly caused, by their χ_2^{post} value.

Practically all switches from MM to SQM result in a χ_2^{post} values around $\pm 180°$ (green dots), meaning that the dihedral angle about the N-O bond is in the trans configuration. By contrast, the four outliers (colored in red) have χ_2^{post} values of approximately 0°. In other words, these switches end in a *cis* configuration around the N-O bond. Thus, for compound **2** PCA not only detects the outliers, but immediately makes clear how these switches differ from the rest.

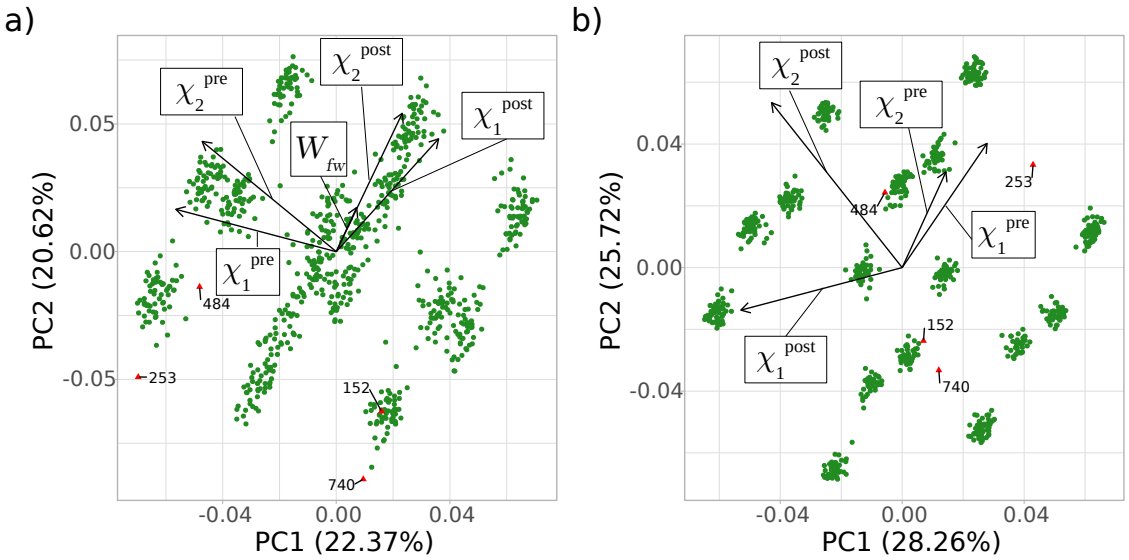


Figure I.9: PCA of compound **7** plotted as PC1 vs. PC2 for the longer equilibration protocol as a) scaled version and b) unscaled version. Outliers are labelled explicitly and marked by red triangles (rather than green dots) according to the predefined outlier criteria (deviations less than $\bar{W} - 3\sigma$ for a), and $|\bar{\chi}_2^{post}| - 3\sigma$ for b), cf. Quasi time series analysis).

Plots for the PCA of compound **7** when using the modified protocol are shown in Figure I.9. The corresponding quasi time series and the PCA for the switches obtained with the original protocol can be found in Figs. A.7 and A.10 of A. Some similarities to **2** are expected as both molecules have an oxime moiety in close proximity to a hydrogen bond acceptor. Figure I.9a shows the plot of the first two principal components for the scaled PCA. Labelling and coloring in red switches based on the $\bar{W} - 3\sigma$ criterion identifies the four outliers which can also be discerned in the quasi time series plot (Figure A.7b in A). As for **2**, these four switches are

located in the opposite direction of the loading vector for W. In this case, the loading vector W has the smallest magnitude of all the variables, thus contributing the least to the variance. On the other hand, the correlation with χ_2^{post} is even more obvious because the corresponding loading vector points almost in the same direction as for W, indicating, that both variables contribute in the same way to the variance (both variables contribute positively to PC1 and PC2). However, the magnitudes of the loading vectors for χ_1, both *pre* and *post*, as well as χ_2^{pre} are quite similar, indicating that all angles contribute equally to the variance. Two of the outliers, which were colored automatically in red, are reasonably well separated (253 and 740), whereas 152 is located in a cluster of "normal" switches and 484 is in close proximity to one. Studying the quasi time series (Figure A.7b) of A), one sees that 152 barely triggers the outlier criterion, which may explain the overlap with other switches. Overall, however, in the case of **7** the outliers are not as well separated as for **2**; the ability to color them automatically is needed.

Figure I.9b shows the first two principle components of the unscaled PCA. Applying the $|\bar{\chi}_2^{post}| - 3\sigma$ criterion identifies all four switches already discerned in Figure I.9a. Analogous to what was observed for **2**, all switches start with χ_2^{pre} around $\pm 180°$, and the four switches ending with χ_2^{post} values around $0°$ result in unusually negative work values. While the outliers (with respect to χ_2^{post}) are somewhat separated from most of the other switches, they would be difficult to discern without the automatic coloring. Nevertheless, PCA facilitated the task considerably. In the scaled PCA, outliers are flagged, i.e., colored in red, based on the $W < \bar{W} - 3\sigma$ criterion, and consideration of the loading vectors strongly indicated to analyze the role of χ_2^{post}. Moreover, it is straightforward to *exclude* other possibilities by applying the $|\chi| < |\bar{\chi}| - 3\sigma$ criterion to the remaining dihedral degrees of freedom. If one does this; very different switches, with work values close to $\bar{W}$, get flagged. This shows that these degrees of freedom are not related to the occurrence of a highly negative work value. Since we use PCA as implemented in R (cf. Methods), these analyses can be carried out interactively and can be automated, at least to some degree. In this manner, such an analysis can be carried out in a short amount of time, even if more than two dihedral degrees of freedom need to be considered.

The analysis of both **2** and **7** indicates a strong correlation between the dihedral angle about the N-O bond after the switch and negative work values ($W < \bar{W} - 3\sigma_W$). As pointed out in earlier work [34, 35], even when using NEW methods configurational mismatches between the different levels of theory are a possible reason for convergence problems. While in this specific case the convergence of the result was not affected

significantly (see Table I.3), the PCA analysis suggests that there are different conformational preferences at the low and high level of theory. A more detailed analysis of configurational sampling about χ_2 at the two levels of theory, i.e., of the production simulations during which restart files were generated, shows that at the MM level of theory the *cis*-configuration about the N-O bond of the oxime moiety present in **2** and **7** is never sampled. By contrast, at the SCC-DFTB level of theory values of $|\chi_2^{pre}| \leq 53°$ were observed $\approx 13\%$ of the time.

The rare forward switches, during which a *trans* $\rightarrow$ *cis* configurational change for χ_2 occurs, lead to untypically negative work values. As just said, χ_2 values around $0°$ are sampled at the SQM level of theory; therefore, such conformational changes during the switch are not artifacts. The remaining question is why they result in more negative work values. In Figure A.2 of A we show a potential energy scan as a function of the χ_1 and χ_2 angles for **7** at the MM level of theory ($\lambda = 0$), the midpoint between MM and SQM ($\lambda = 0.5$) and the SQM level of theory ($\lambda = 1$). At the SQM endpoint, χ_2 values of $\pm 180°$ and $0°$ are quite close energetically, although $\chi_2 = \pm 180°$ is the global minimum conformation. The comparison of the three energy surfaces shown in Figure A.2 of A suggests the following explanation: As the potential energy surface changes from MM to SQM, a shallow minimum around $\chi_1 = \pm 180°/\chi_2 = 0°$ develops, which is already clearly discernible in the plot for $\lambda = 0.5$ (Figure A.2of A, middle). At $\lambda = 0.5$ the barrier separating the minima at $\chi_1 = \pm 180°/\pm\chi_2 = 180°$ and $\chi_1 = \pm 180°/\chi_2 = 0°$ is lower by about 2kcal/mol than at the SQM endpoint (≈ 4 instead of > 6 kcal/mol). Therefore, conformational changes from $\chi_2 = \pm 180°$ to $0°$ can certainly occur, although given the short switching length of 2 ps, it is not surprising that they are rare. As one sees from Figure A.2 (left) in A, at the MM level of theory the minimum energy basin about $\chi_1/\chi_2 = \pm 180°/\pm 180°$ is quite wide, particularly in the χ_2 direction. Thus, χ_2 values of $\pm 100°$ or even $\pm 90°$ are quite accessible and are sampled at the MM level of theory. In other words, starting configurations $|\chi_2| < 100°$ are close to the top of the barrier at $\lambda = 0.5$ and if their starting velocities point in the right direction, the barrier can be overcome. Indeed, all four switches that ended in the $\chi_1 = \pm 180°/\chi_2 = 0°$ minimum started from such configurations. As the switch progresses, the system then slides down towards the local energy minimum at $\chi_2 = 0°$ at the SQM endpoint, resulting in a more negative work value compared to systems which remain around $\chi_1/\chi_2 = \pm 180°/\pm 180°$ throughout the switching simulation.

I.4 Discussion/Conclusions

We systematically investigated the convergence of Jarzynski's equation applied to computing $\Delta A^{MM \to SQM}$ as a function of switching length and number of non-equilibrium work values used for averaging. The data clearly indicate that using fewer, but longer switches leads to better converged and, hence, more accurate results. This is in line with earlier results [36], but it was important to ascertain that this does hold even in the regime of very short switching lengths, given that all the protocols which are computationally feasible, including NSWI2000, are far from equilibrium. Our findings lead to immediate practical benefits. Using the best protocol (800 switches of 2 ps length, NSWI2000) reduces the computational cost compared to the naive protocol of the original "HiPen" study [35] by almost a factor of 10. Having identified the NSWI2000 protocol as a reliable way to compute $\Delta A^{low \to high}$, we optimized it even further, 200 rather than 800 switches ($\text{NSWI2000}^{\text{red}}$). With this final protocol we obtained as good if not better results compared to Ref. [35] at one fortieth of the cost. At the same time, the present protocol remains trivially parallel and if multiple computers and/or CPU cores are available, one can compute the corrections $\Delta A^{MM \to SQM}$ efficiently.

When attempting to understand possible sources of poor or slow convergence, two tools proved very helpful. (i) Plotting the work values not only as a histogram, but as a quasi time series, and (ii) employing PCA to detect correlations between outliers with respect to work values and conformational degrees of freedom before and after the switch. It should be stressed that these tools/utilities require as input either quantities one needs to compute anyways, such as the non-equilibrium work values, or quantities which can be calculated extremely fast, such as dihedral angles. In other words, very useful insights can be obtained at little or no cost. We employed both methods using the statistical software system R, which provides a convenient interface for plotting and carrying out the PCA. This facilitated detecting and understanding the source of outliers considerably. However, once the raw data are available, various programs could be used for these analyses.

In all cases which we analyzed in detail, the cause of outlying work values could be traced to a conformational degree of freedom, which, for these relatively simple systems, was always a dihedral angle. In many applications of computational chemistry, rotamers need to be enumerated and/or the corresponding dihedral angles identified. Our experience suggests that such degrees of freedom should be routinely analyzed, e.g. by PCA, to detect and understand possible convergence problems.

In this study we deliberately focused on the "good" compounds from the full "HiPen" data set. More efficient protocols to obtain the work values needed for

Jarzynski's equation may not be enough for the "bad" and "ugly" cases. Here, techniques like force matching [68] to make the low level of theory more high level like or the judicious use of intermediate stages to carry out the low to high transformation [31] may be needed. Further, better equilibration and sampling strategies, such as self-guided Langevin dynamics (SGLD) [69, 70], in preparation for the equilibrium simulations during which the staring points for the switching step are written, should prove helpful. The combination with optimized protocols to obtain the non-equilibrium work values needed will make it possible to compute the correction $\Delta A^{MM \to SQM}$ reliably and with sufficient efficiency.

I.5 Abbreviations

MD	Molecular Dynamics
MM	Molecular Mechanics
(S)QM	(Semi-empirical-)Quantum Mechanics
(S)QM/MM	(Semi-empirical-)Quantum Mechanical/Molecular Mechanical hybrid methods
FES	Free energy simulation
NEW	Non-equilibrium work methods
JAR	Jarzynski's Equation
CRO	Crooks' Equation

Supporting Information

The following can be found in supporting information. Table A.1: additional details of model compounds; Table A.2: summary statistics for simulations with the short equilibration protocol, Tables A.3–A.6: detailed results for data plotted in Figures I.3 and I.7, as well as Figures A.3–A.5. Figure A.1: change in dihedral angle labelling for compound **7**; Figure A.2: Potential energy scan about dihedral angles χ_1 and χ_2 for **7**; Figures A.3–A.5: alternative plots of results; Figures A.6, A.7: quasi time series analysis for **2** and **7**; Figure A.8: convergence of the results as a function of the switching length for **12**, **2**, **7**, **10** and **21** when using the extended equilibration protocol; Figures A.9 and A.10: PCA of data for **2** and **7** generated with the initial, short equilibration protocol.

Acknowledgement

The work was supported by grants NIH 1R01GM129519-01 (HLW) and FWF P-31024 (SB). The authors thank Phillip Hudson for technical help and critical comments on an early draft of the manuscript.

Bibliography

[1] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten, "Discovery through the computational microscope," *Structure*, vol. 17, no. 10, pp. 1295–1306, 2009.

[2] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, "Molecular dynamics simulations of large macromolecular complexes," *Curr. Opin. Struct. Biol.*, vol. 31, pp. 64–74, 2015.

[3] Y. Miao and J. A. McCammon, "Unconstrained enhanced sampling for free energy calculations of biomolecules: a review," *Mol. Simul.*, vol. 42, no. 13, pp. 1046–1055, 2016.

[4] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren, "Basic ingredients of free energy calculations: a review," *J. Comput. Chem.*, vol. 31, no. 8, pp. 1569–1582, 2010.

[5] E. Vanden-Eijnden, "Some recent techniques for free energy calculations," *J. Comput. Chem.*, vol. 30, no. 11, pp. 1737–1747, 2009.

[6] Z. Cournia, B. Allen, and W. Sherman, "Relative binding free energy calculations in drug discovery: Recent advances and practical considerations," *J. Chem. Inf. Model.*, vol. 57, pp. 2911–2937, dec 2017.

[7] M. Aldeghi, J. P. Bluck, and P. C. Biggin, "Absolute alchemical free energy calculations for ligand binding: A beginner's guide," in *Methods in Molecular Biology*, pp. 199–232, Springer New York, 2018.

[8] R. Abel, L. Wang, D. L. Mobley, and R. A. Friesner, "A critical review of validation, blind testing, and real- world use of alchemical protein-ligand binding free energy calculations," *Curr. Top. Med. Chem.*, vol. 17, aug 2017.

[9] Y. Wang, C. Peng, G. Wang, Z. Xu, Y. Luo, J. Wang, and W. Zhu, "Exploring binding mechanisms of VEGFR2 with three drugs lenvatinib, sorafenib, and sunitinib by molecular dynamics simulation and free energy calculation," *Chem. Biol. Drug Des.*, vol. 93, no. 5, pp. 934–948, 2019.

[10] L. C. James and D. S. Tawfik, "Conformational diversity and protein evolution–a 60-year-old hypothesis revisited," *Trends Biochem. Sci.*, vol. 28, no. 7, pp. 361–368, 2003.

[11] Q. Li, L. Xiong, J. Gao, and H.-Y. Zhang, "Molecular mechanisms of the protein-protein interaction–regulated binding specificity of basic-region leucine zipper transcription factors," *J. Mol. Model.*, vol. 25, no. 8, pp. 1–11, 2019.

[12] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, "Atomic-level characterization of protein–protein association," *Proc. Nat. Acad. Sc.*, vol. 116, no. 10, pp. 4244–4249, 2019.

[13] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *J. Mol. Biol.*, vol. 103, no. 2, pp. 227–249, 1976.

[14] U. C. Singh and P. A. Kollman, "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the ch3cl+ cl- exchange reaction and gas phase protonation of polyethers," *J. Comput. Chem.*, vol. 7, no. 6, pp. 718–730, 1986.

[15] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations," *J. Comput. Chem.*, vol. 11, no. 6, pp. 700–733, 1990.

[16] F. Maseras and K. Morokuma, "IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states," *J. Comput. Chem.*, vol. 16, no. 9, pp. 1170–1179, 1995.

[17] J. C. Smith and B. Roux, "Eppur si muove! The 2013 Nobel Prize in Chemistry," *Structure*, vol. 21, no. 12, pp. 2102–2105, 2013.

[18] D. Berta, P. J. Buigues, M. Badaoui, and E. Rosta, "Cations in motion: QM/MM studies of the dynamic and electrostatic roles of H+ and Mg2+ ions in enzyme reactions," *Curr. Opin. Struct. Biol.*, vol. 61, pp. 198–206, 2020.

[19] H. J. Kulik, "Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer," *Phys. Chem. Chem. Phys.*, vol. 20, no. 31, pp. 20650–20660, 2018.

[20] V. Vennelakanti, A. Nazemi, R. Mehmood, A. H. Steeves, and H. J. Kulik, "Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins," *Curr. Opin. Struct. Biol.*, vol. 72, pp. 9–17, 2022.

[21] H. S. Fernandes, M. J. Ramos, and N. M. Cerqueira, "Catalytic mechanism of the serine hydroxymethyltransferase: a computational ONIOM QM/MM study," *ACS Catal.*, vol. 8, no. 11, pp. 10096–10110, 2018.

[22] M. Liu, Y. Wang, Y. Chen, M. J. Field, and J. Gao, "QM/MM through the 1990s: The first twenty years of method development and applications," *Isr. J. Chem.*, vol. 54, no. 8-9, pp. 1250–1263, 2014.

[23] Q. Cui, T. Pal, and L. Xie, "Biomolecular qm/mm simulations: What are some of the "burning issues"?," *J. Phys. Chem. B*, vol. 125, no. 3, pp. 689–702, 2021.

[24] W. Yang, Q. Cui, D. Min, and H. Li, "QM/MM Alchemical Free Energy Simulations: Challenges and Recent Developments," *Annu. Rep. Comput. Chem.*, vol. 6, pp. 51–62, 2010.

[25] X. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov, and Q. Cui, "QM/MM Free Energy Simulations: Recent Progress and Challenges," *Mol. Simul.*, vol. 42, no. 13, pp. 1056–1078, 2016.

[26] J. Heimdal and U. Ryde, "Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations," *Phys. Chem. Chem. Phys.*, vol. 14, no. 36, pp. 12592–12604, 2012.

[27] S. Boresch and H. L. Woodcock, "Convergence of single-step free energy perturbation," *Mol. Phys.*, vol. 115, no. 9-12, pp. 1200–1213, 2017.

[28] G. König, P. S. Hudson, S. Boresch, and H. L. Woodcock, "Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-boltzmann bennett reweighting schemes," *J Chem Theory Comput*, vol. 10, pp. 1406–1419, mar 2014.

[29] G. König, B. R. Brooks, W. Thiel, and D. M. York, "On the convergence of multi-scale free energy simulations," *Mol Simulat*, vol. 44, pp. 1062–1081, may 2018.

[30] T. J. Giese and D. M. York, "Development of a robust indirect approach for MM $\rightarrow$ QM free energy calculations that combines force-matched reference potential and bennett's acceptance ratio methods," *J Chem Theory Comput*, vol. 15, pp. 5543–5562, sep 2019.

[31] S. Ito and Q. Cui, "Multi-level free energy simulation with a staged transformation approach," *J. Chem. Phys.*, vol. 153, p. 044115, jul 2020.

[32] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, no. 14, p. 2690, 1997.

[33] P. S. Hudson, H. L. Woodcock, and S. Boresch, "Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems," *J. Phys. Chem. Lett.*, vol. 6, no. 23, pp. 4850–4856, 2015.

[34] F. L. Kearns, P. S. Hudson, H. L. Woodcock, and S. Boresch, "Computing Converged Free Energy Differences Between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1376–1388, 2017.

[35] F. L. Kearns, L. Warrensford, S. Boresch, and H. L. Woodcock, "The Good, the Bad, and the Ugly: "HiPen", a New Dataset for Validating (S)QM/MM Free Energy Simulations," *Molecules*, vol. 24, no. 4, p. 681, 2019.

[36] G. Hummer, "Fast-growth thermodynamic integration: Error and efficiency analysis," *J. Chem. Phys.*, vol. 114, pp. 7330–7337, may 2001.

[37] D. Wu and D. A. Kofke, "Asymmetric bias in free-energy perturbation measurements using two hamiltonian-based models," *Phys Rev E*, vol. 70, p. 066702, dec 2004.

[38] A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free-energy calculations," *J. Phys. Chem. B*, vol. 114, pp. 10235–10253, jul 2010.

[39] C. Dellago and G. Hummer, "Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics," *Entropy*, vol. 16, no. 1, pp. 41–61, 2013.

[40] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, no. 3, p. 2361, 2000.

[41] T. Schmiedl and U. Seifert, "Optimal finite-time processes in stochastic thermodynamics," *Phys. Rev. Lett.*, vol. 98, no. 10, p. 108301, 2007.

[42] H. Then and A. Engel, "Computing the optimal protocol for finite-time processes in stochastic thermodynamics," *Phys. Rev. E*, vol. 77, no. 4, p. 041105, 2008.

[43] P. Geiger and C. Dellago, "Optimum protocol for fast-switching free-energy calculations," *Phys. Rev. E*, vol. 81, no. 2, p. 021127, 2010.

[44] M. Aldeghi, V. Gapsys, and B. L. de Groot, "Accurate estimation of ligand binding affinity changes upon protein mutation," *ACS Cent. Sci.*, vol. 4, pp. 1708–1718, dec 2018.

[45] Maybridge HitFinder^{TM} (14.400 compounds). `https://www.maybridge.com/portal/alias_Rainbow/lang_en/tabID_229/DesktopDefault.aspx`, 2020. Accessed: 2020-09-11.

[46] K. Vanommeslaeghe and A. D. MacKerell Jr, "Automation of the charmm general force field (cgenff) i: bond perception and atom typing," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3144–3154, 2012.

[47] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell Jr, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3155–3168, 2012.

[48] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

[49] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Phys. Rev. B*, vol. 58, no. 11, p. 7260, 1998.

[50] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, "A QM/MM implementation of the self-consistent charge density functional tight binding (scc-dftb) method," *J. Phys. Chem. B*, vol. 105, no. 2, pp. 569–585, 2001.

[51] M. A. Olsson and U. Ryde, "Comparison of QM/MM Methods to Obtain Ligand-Binding Free Energies," *J. Chem. Theory Comput.*, vol. 13, pp. 2245–2253, Mar. 2017.

[52] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, "CHARMM: the biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.

[53] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[54] J.-W. Chu, B. L. Trout, and B. R. Brooks, "A super-linear minimization scheme for the nudged elastic band method," *J. Chem. Phys.*, vol. 119, no. 24, pp. 12708–12717, 2003.

[55] S. Nsikabaka, W. Harb, and M. Ruiz-López, "The role of water on the acid-promoted E/Z isomerization of oximes in aqueous solution," *J. Mol. Struct. THEOCHEM*, vol. 764, no. 1-3, pp. 161–166, 2006.

[56] M. Gaus, A. Goez, and M. Elstner, "Parametrization and Benchmark of DFTB3 for Organic Molecules," *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 338–354, 2013. PMID: 26589037.

[57] M. Gaus, X. Lu, M. Elstner, and Q. Cui, "Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1518–1537, 2014. PMID: 24803865.

[58] X. Lu, M. Gaus, M. Elstner, and Q. Cui, "Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 1062–1082, 2015. PMID: 25178644.

[59] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, "Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems," *J. Chem. Theory Comput.*, vol. 11, no. 1, pp. 332–342, 2015.

[60] H. L. Woodcock, B. T. Miller, M. Hodoscek, A. Okur, J. D. Larkin, J. W. Ponder, and B. R. Brooks, "Mscale: a general utility for multiscale modeling," *J. Chem. Theory Comput.*, vol. 7, no. 4, pp. 1208–1219, 2011.

[61] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[62] R. Becker, J. Chambers, and A. Wilks, "The new s language. wadsworth and brooks/cole computer science series," *Pacific Grove*, 1988.

[63] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 2015.

[64] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *J. R. Stat. Soc., B: Stat. Methodol.*, vol. 53, no. 3, pp. 683–690, 1991.

[65] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[66] W. Venables and B. Ripley, "Modern applied statistics with s springer verlag," *New York*, 2002.

[67] K. V. Mardia, J. Kent, and J. Bibby, *Multivariate analysis*. London: Academic Press, 1979.

[68] P. S. Hudson, S. Boresch, D. M. Rogers, and H. L. Woodcock, "Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching," *J. Chem. Theory Comput.*, vol. 14, no. 12, pp. 6327–6335, 2018.

[69] X. Wu and B. R. Brooks, "Self-guided langevin dynamics simulation method," *Chem. Phys. Lett.*, vol. 381, no. 3-4, pp. 512–518, 2003.

[70] G. König, B. T. Miller, S. Boresch, X. Wu, and B. R. Brooks, "Enhanced sampling in free energy calculations: Combining sgld with the bennett's acceptance ratio and enveloping distribution sampling methods," *J. Chem. Theory Comput.*, vol. 8, no. 10, pp. 3650–3662, 2012. PMID: 26593010.

[71] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman, "ZINC: A Free Tool to Discover Chemistry for Biology," *J. Chem. Inf. Model.*, vol. 52, no. 7, pp. 1757–1768, 2012. PMID: 22587354.

[72] R. H. Wood, W. C. F. Muhlbauer, and P. T. Thompson, "Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis," *J. Phys. Chem.*, vol. 95, no. 17, pp. 6670–6675, 1991.

[73] D. Wu and D. A. Kofke, "Model for small-sample bias of free-energy calculations applied to gaussian-distributed nonequilibrium work measurements," *J. Chem. Phys.*, vol. 121, no. 18, pp. 8742–8747, 2004.

II. Exploring Routes to Enhance the Calculation of Free Energy Differences via Nonequilibrium Work SQM/MM Switching Simulations Using Hybrid Charge Intermediates between MM and SQM Levels of Theory or Non-Linear Switching Schemes

Contents

Bibliographic Information

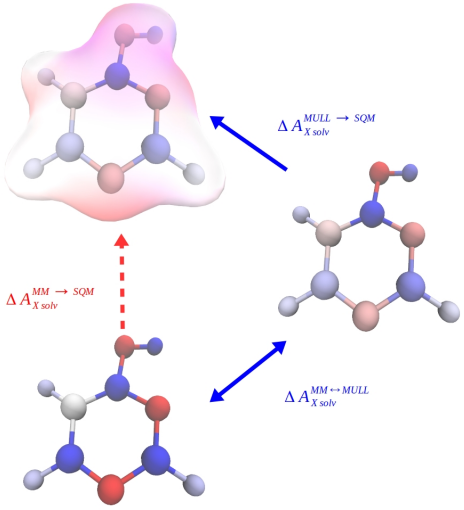
Schöller, A. Woodcock,H. L., Boresch, S. Exploring Routes to Enhance the Calculation of Free Energy Differences via Nonequilibrium Work SQM/MM Switching Simulations Using Hybrid Charge Intermediates between MM and SQM Levels of Theory or Non-Linear Switching Schemes *Molecules* **2023** *28*, 4006. MDPI. DOI: 10.3390/molecules28104006.

Author's contribution

Conceptualization, A.S., H.L.W. and S.B.; methodology, H.L.W. and S.B.; validation, A.S. and S.B.; formal analysis, A.S.; investigation, A.S.; resources, S.B.; writing—original draft preparation, A.S.; writing—review and editing, H.L.W. and S.B.; visualization, A.S.; supervision, S.B.; project administration, S.B.; funding acquisition, H.L.W. and S.B. All authors have read and agreed to the published version of the manuscript.

Copyright Notice

Abstract



Non-equilibrium work switching simulations and Jarzynski's equation are a reliable method for computing free energy differences, $\Delta A^{\text{low}\rightarrow\text{high}}$, between two levels of theory, such as a pure molecular mechanical (MM) and a quantum mechanical/molecular mechanical (QM/MM) description of a system of interest. Despite the inherent parallelism, the computational cost of this approach can quickly become very high. This is particularly true for systems where the core region, the part of the system to be described at different levels of theory, is embedded in an environment such as explicit solvent water. We find that even for relatively simple solute–water systems, switching lengths of at least 5 ps are necessary to compute $\Delta A^{\text{low}\rightarrow\text{high}}$ reliably. In this study, we investigate two approaches towards an affordable protocol, with an emphasis on keeping the switching length well below 5 ps. Inserting a hybrid charge intermediate state with modified partial charges, which resembles the charge distribution of the desired high level, makes it possible to obtain reliable calculations with 2 ps switches. Attempts using step-wise linear switching paths, on the other hand, did not lead to improvement, i.e., a faster convergence for all systems. To understand these findings, we analyzed the solutes' properties as a function of the partial charges used and the number of water molecules in direct contact with the solute, and studied the time needed for water molecules to reorient themselves upon a change in the solute's charge distribution.

II.1 Introduction

Since the first applications reported shortly before 1990 [1, 2], so-called alchemical free energy simulations (FESs) have become an essential tool of computational chemists in academia and industry alike [3, 4, 5, 6]. The free energy difference determines the spontaneities of chemical or biochemical processes, and hence, makes it possible to predict, e.g., the binding affinities of potential drugs [7, 8, 9]. In addition to computationally estimating an important macroscopic, thermodynamic quantity, FESs can also provide insight in the microscopic origins [10]. One source of error is the accuracy with which interactions within and between molecules are modeled in the underlying molecular dynamics simulations. In certain applications, force field-based descriptions may be insufficient [11]. Even if one is not interested in studying chemical reactions, the neglect of induced electronic polarization in classical force fields may lead to problems, even for seemingly simple applications, such as the computation of relative free energies of hydration of mono- or bivalent ions [12, 13, 14]. In such situations, hybrid quantum-mechanical/molecular mechanical (QM/MM) Hamiltonians are called for. However, even when using only semi-empirical quantum chemical methods (henceforth abbreviated as SQMs) for the core region, i.e., the part of the system that one is particularly interested in, the computational cost can quickly become prohibitive. Further, several tricks that are frequently used in alchemical FES do not work with SQM/MM Hamiltonians [15]. Early on, Warshel, Gao, and others pioneered the use of indirect cycles, also referred to as "multilevel" free energy simulations, as depicted in Figure II.1a), which have become a widely used way of performing FES with an (S)QM/MM description of interactions [16, 17, 18].

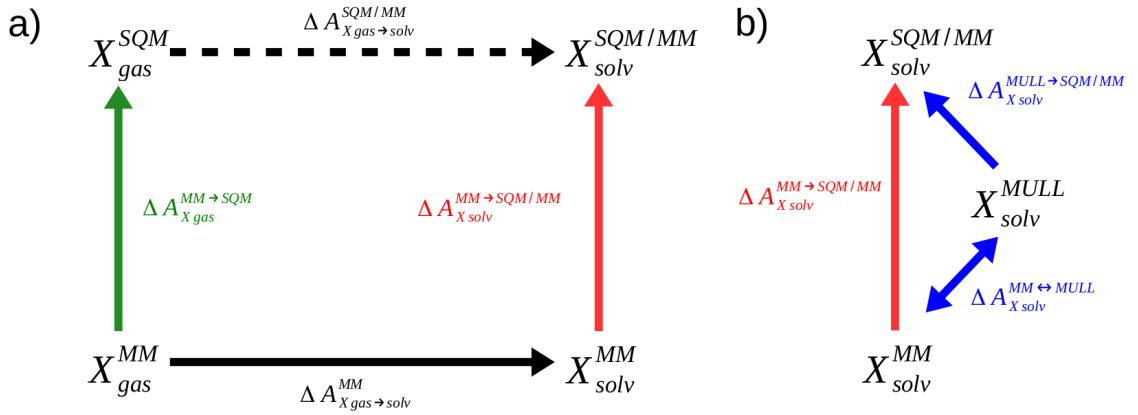


Figure II.1: (**a**) Indirect alchemical thermodynamic cycle to compute $\Delta A^{SQM/MM}_{X\,gas\to solv}$ (dashed arrow) in three steps according to Equation (II.1). (**b**) Indirect alchemical thermodynamic cycle to compute $\Delta A^{MM\to SQM/MM}_{X\,solv}$ (red arrow) via hybrid charge intermediates; cf. Equation (II.8).

To illustrate the indirect approach, we consider the calculation of the aqueous solvation free energy of a solute X. As can be seen in Figure II.1a, the quantity of interest and the solvation free energy difference $\Delta A^{SQM/MM}_{X\ gas\rightarrow solv}$ at the high SQM/MM level of theory (dashed, black arrow in Figure II.1a), can also be obtained in three steps according to

$$\Delta A^{SQM/MM}_{X\ gas\rightarrow solv} = -\Delta A^{MM\rightarrow SQM}_{X\ gas} + \Delta A^{MM}_{X\ gas\rightarrow solv} + \Delta A^{MM\rightarrow SQM/MM}_{X\ solv} \tag{II.1}$$

In addition to computing the solvation free energy $\Delta A^{MM}_{X\ gas\rightarrow solv}$ at the force field (MM) level of theory (solid black arrow), two correction steps, accounting for the free energy difference of X in the gas phase and the aqueous solution between the two levels of theory (green and red arrows), are required. The development in this field over recent years shows that there is not only enormous potential, but also a high degree of interest in the FES community to use this strategy to compute free energy differences at the SQM/MM levels of theory [19, 20, 21, 22, 23, 24, 25, 26, 27].

Work by Heimdal and Ryde [28] made clear that the challenging step of indirect cycle SQM/MM FES is the calculation of the low-to-high corrections $\Delta A^{MM\rightarrow SQM}_{X\ gas}$ and $\Delta A^{MM\rightarrow SQM/MM}_{X\ solv}$: the green and red arrows of Figure II.1a. Various approaches have been used, often requiring a sophisticated decomposition of the cycle and the employment of fitting procedures in order to be successful [18, 29]. In the past, we showed that the corrections $\Delta A^{MM\rightarrow SQM}$ can be computed reliably and accurately using non-equilibrium work (NEW) methods [30, 31, 32, 33, 34, 35]. In our work to date, computationally expensive protocols were used, which are unsuitable for most real-world applications. Recently, we optimized the NEW protocols regarding switching length and the number of switches, and presented a sufficiently efficient workflow for practical applications [36]. However, these optimized protocols were tested so far only in the gas phase, i.e., the green arrow of Figure II.1a ($\Delta A^{MM\rightarrow SQM}_{X\ gas}$).

Here, we investigate whether the protocols suggested in Ref. [36] are also suitable for aqueous solution, i.e., when the region of interest is to be treated at a high level of theory, e.g., SQM, interacts with an MM environment ($\Delta A^{MM\rightarrow SQM/MM}_{X\ solv}$, red arrow in Figure II.1a). In earlier work, we had noted that NEW switching simulations from a pure MM to an SQM/MM level of theory converged more slowly compared to gas phase transformations involving only MM to SQM, and observed that the charge distribution resulting from the fixed charges of the force field and the average charge distribution obtained at the SQM/MM level of theory were noticeably different [31]. We speculated that the necessary reorientation of water in response to such a change in the charge distribution of the solute might be responsible for the slower convergence. Using equilibrium methods to compute $\Delta A^{MM\rightarrow SQM/MM}_{X\ solv}$, Ito and Cui

reported that inserting an intermediate state with fixed charges more similar to the charge distribution at the SQM/MM target state improved convergence considerably [37].

The solvent-reorientation dynamics of water in response to charge changes is experimentally well characterized and understood [38, 39]. Two relaxation times, one faster than 0.5 ps and the other approximately 2 ps, can be discerned. Therefore, if the charge distribution of the region of interest is significantly different at the two levels of theory, NEW switching simulations of only 2 ps, as used in Ref. [31], may be too short to obtain converged results. With this in mind, we tested two strategies. (i) Similarly to the work by Ito and Cui, we introduce an intermediate state, which is still purely MM, but with the atomic partial charges modified to resemble the average charge distribution of the SQM description at the high level of theory. (ii) Since water reorientation in response to a change in charge distribution occurs on two timescales, the linear switching protocols we have used so far may be suboptimal. In particular, we test whether protocols in which switching is carried out initially at a faster rate, e.g., switching from the MM to a 35% SQM/MM description at just 10% of the switching length, followed by a second slower phase during the remaining 90% of the switching length, might lead to improved convergence. Obviously, both strategies can be combined. Model calculations exploring these two approaches are complemented by a series of analyses in which we investigate the effects of a solute's charge distribution on properties such as its dipole moment, as well as on the interactions with surrounding waters.

The focus of this work is on the effect by which the charge distribution of a solute described at different levels of theory has on the convergence of calculating $\Delta A_{X\ solv}^{MM \rightarrow SQM/MM}$. To avoid unrelated complications from too-flexible molecules, where conformational preferences might be different at the low and high levels of theory, respectively, we mostly chose rather rigid model compounds for the test calculations. Rather than using (a subset of) the so-called "HiPen" test set [34], we selected tautomeric pairs (cf. Methods). Most of the compounds are rather rigid without rotatable bonds. Therefore, the pure MM and the SQM/MM descriptions of the systems—the solute (tautomer) are described by SQM, and all waters remain classical—differing primarily by the charge distributions of the solute at the two levels of theory.

The remainder of the manuscript is organized as follows. In Results, we first study the convergence of NEW switching simulations, using the optimized protocol from our preceding study [36] (switching lengths of 2 ps, 200 NEW switches per transformation). We then investigate how convergence $\Delta A_{X\ solv}^{MM \rightarrow SQM/MM}$ improves

when employing switching lengths of 5 ps, which, however, may be too costly for many applications. Next, we present results using the two mitigation strategies outlined above. Our findings are rationalized using additional data characterizing the differences in solute properties, in particular the dipole moment at the two levels of theory, as well as details on the solvent-reorientation dynamics of water under the simulation conditions. In Section II.4, we summarize the theoretical background, introduce the model systems, and provide the technical details of all simulations and analyses carried out.

II.2 Results

II.2.1 Overview of Calculated Free Energy Differences and Paths

Below, several approaches to calculate the free energy difference between two levels of theory are compared. To specify the method and the exact meaning of a free energy difference, we adopt the following labeling scheme:

$$\Delta A_{Xphase}^{level\ of\ theory\ I \rightarrow/\leftrightarrow\ level\ of\ theoy\ II}$$

Here, the subscript X indicates the compound; the qualifier *phase* can either be "gas" (gas phase) or "solv" (solvated phase). In the superscript, *level of theory I* and *level of theory II* can be "MM", "MULL(gas)", "MULL(solv)", "MULL(solv*)", or "SQM(/MM)". MM indicates the use of the regular force field. The abbreviation SQM indicates that the SCC-DFTB semi-empirical method was used to compute interactions, either for the full system (the gas phase), or for the solute (the aqueous solution; waters were always described classically). The labels MULL(gas), MULL(solv), and MULL(solv*) refer to three methods used to derive the average Mulliken charges described in Section II.4.3.

A simple arrow $\rightarrow$ stands for a one-sided method to compute the free energy difference; in this work, this always means Jarzynki's equation (JAR) [40]. The double-pointed arrow $\leftrightarrow$ indicates the use of a two-sided method. As described in Section II.4, we used Crooks' equation (CRO) [41] to compute the reference results. Furthermore, the free energy differences between MM and the intermediate state with modified partial charges, i.e., MULL(gas), MULL(solv), or MULL(solv*), were computed using Bennett's acceptance ratio method (BAR) [42].

As in our previous work, all results are reported relative to a reference result; i.e., instead of reporting, e.g., $\Delta A_{Xphase}^{MM\rightarrow SQM}$, we report the corresponding double free

energy difference $\delta\Delta A_{Xphase} = \Delta A_{Xphase}^{MM \rightarrow SQM} - \Delta A_{Xphase}^{\mathrm{CRO_{Ref}}}$. In the past, the reference results $\Delta A_{Xphase}^{\mathrm{CRO_{Ref}}}$ were obtained with CRO, based on 200 instances of 2 ps forward and backward NEW switches. As described in Section II.2.2 below, we decided to use the CRO results obtained from 5 ps switches as the reference ($\mathrm{CRO_{Ref}}$). The exact definition of $\delta\Delta A_{Xphase}$ depends on whether the free energy difference between two levels of theory is computed in a single step, i.e.,

$$\delta\Delta A_{Xphase} = \Delta A_{Xphase}^{\mathrm{MM} \rightarrow/\leftrightarrow \mathrm{SQM}} - \Delta A_{Xphase}^{\mathrm{CRO_{Ref}}} \tag{II.2a}$$

or via an hybrid charge intermediate state. Here, we have

$$\delta\Delta A_{Xphase} = \Delta A_{Xphase}^{\mathrm{MM} \leftrightarrow \mathrm{MULL}} + \Delta A_{Xphase}^{\mathrm{MULL} \rightarrow \mathrm{SQM}} - \Delta A_{Xphase}^{\mathrm{CRO_{Ref}}} \tag{II.2b}$$

to account for the correction step between the MM representation and the hybrid Mulliken charge intermediate state; cf. Equation (II.8).

II.2.2 Performances of 2 and 5 ps Linear Switching Protocols in Solution

Before carrying out calculations in aqueous solution, the optimized JAR protocol from Ref. [36] was tested for all compounds in the gas phase. All results can be found in Table B.3 of B. The model systems used in this work are described in Section II.4.2. Except for **1-t1** (see further down), $|\delta\Delta A|$ was < 0.1 kcal/mol. While $\delta\Delta A(\mathbf{1\text{-}t1}) \approx 0.35$ kcal/mol is slightly larger than the ideal maximum deviation of ± 0.25 kcal/mol, this value is still acceptable for most practical applications. Thus, the protocol recommendations of our previous study [36] are applicable to the model compounds studied here. Based on the gas phase results, any noticeable deterioration of $\delta\Delta A$ therefore has to be caused by differences in the description of solute–solvent interactions at the two levels of theory.

In Figure II.2, the $\delta\Delta A$ values for all compounds obtained via the three protocols/methods are compared against the reference results obtained from 200 forward/backward switches of 5 ps length: 200 MM to SQM switches of 2 ps length, the recommended protocol from Ref. [36], (JAR(2ps), red circles), 200 MM to SQM switches of 5 ps length (JAR(5ps), orange circles), and the CRO results obtained from 200 forward/backward switches of 2 ps length (CRO(2ps), blue circles). The data visualized in Figure II.2 are tabulated in Table B.2; even more detailed results can be found in Tables B.4 and B.5. For the shortest protocol, JAR(2ps), four results lie outside the ± 0.25 kcal/mol threshold indicated by the two dashed lines in the

figure. Taking error bars into account (see Table B.2), the results for **1-t1** and **8-t2** can be considered as acceptable, $|\delta\Delta A| < 0.4$ kcal/mole; furthermore, $\delta\Delta A$ (**1-t1**) is comparable to the value obtained in the gas phase (see Table B.3). However, for **5-t2** and **6-t2**, the deviation from the reference result is larger than 1 kcal/mol and would lead to a sizable systematic error.

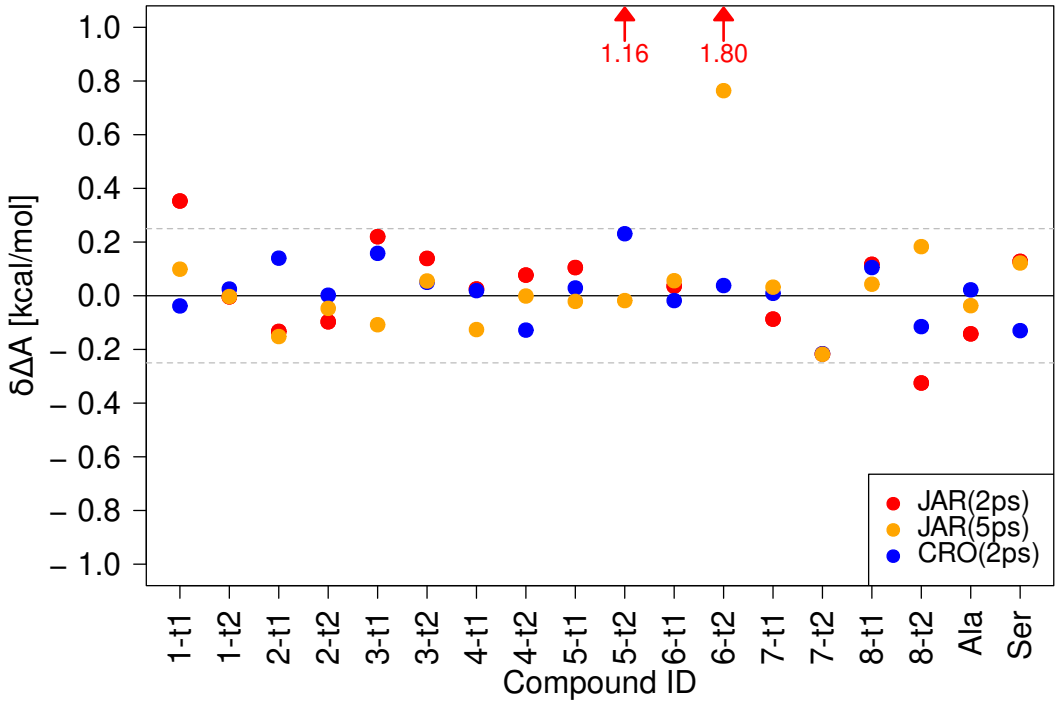


Figure II.2: Comparison of $\delta\Delta A^{MM\to SQM}_{Xsolv}$ using three protocols without intermediate state: JAR(2ps): 2 ps forward switches from MM to SQM, JAR(5ps): 5 ps forward switches from MM to SQM, and CRO(2ps): 2 ps forward and backward switches. In all cases, the CRO results obtained from 5 ps switches were considered as the reference values $\mathrm{CRO_{Ref}}$. For **5t-2** and **6t-2**, the JAR(2ps) results are off-scale, which is indicated by the red arrows.

The use of a 5 ps switching length (JAR(5ps), orange circles) certainly helps, but one poor result, **6-t2**, $\delta\Delta A \approx 0.8$ kcal/mol, remains. Thus, the relatively inexpensive protocol suggested in Ref. [36] cannot be recommended for calculations in solution. Even worse, the use of more costly switching simulations of 5 ps length does not help in all cases.

Figure II.2 also shows results for the CRO(2ps) protocol (blue circles). Compared to our findings in the gas phase [36], the differences to the reference result CRO(5ps) are noticeably more pronounced. Furthermore, as can be seen in Tables B.4 and B.5, the statistical uncertainty of the CRO(2ps) results was consistently higher than those of the CRO(5ps) results, even though the overlap between the forward and backward work distributions seemed adequate in all cases. While all CRO(2ps) $\delta\Delta A$

values lie within the ±0.25 kcal/mol threshold, we nevertheless decided to choose CRO(5ps) as the reference protocol.

Three of the four compounds failing the ±0.25 kcal/mol threshold using the JAR(2ps) protocol, **5-t2**, **6-t2**, and **8-t2**, are lactams. Interestingly, the corresponding tautomeric lactim states do not cause any problems. The pair **1-t1**/**1-t2** belongs to the keto-enol class of tautomerism. For **1-t1**, $\delta\Delta A \approx 0.35$ kcal/mol in both the gas phase and in aqueous solution; therefore, different conformational preferences at the two levels of theory may be responsible for the poor convergence when using switching lengths of 2 ps.

II.2.3 Performances of Hybrid Charge Intermediates

In Figure II.3, we summarize the performances of the indirect NEW switching protocols using a hybrid intermediate charge state in terms of $\delta\Delta A$, cf. Equation (II.2b). As described in Methods (Section II.4.3), three approaches to obtain average Mulliken charges were tested: averaging over gas phase configurations (MULL(gas), over configurations sampled in aqueous solution at the MM level of theory (MULL(solv)), and over configurations sampled in aqueous solution at the SQM/MM level of theory (MULL(solv*)). The results obtained with the direct JAR(2ps) protocol (red circles), already shown in Figure II.2 and henceforth referred to as just MM, are included as well.

Clearly, all three protocols employing an intermediate hybrid charge state perform much better than the direct NEW switches (MM), though not to the same degree. Using the MULL(gas) intermediate state (orange triangles), the ±0.25 kcal/mol threshold for **5-t2** is still missed, $\delta\Delta A = 0.32$ kcal/mol, and the result for **Ala** becomes worse ($\delta\Delta A = -0.49$ kcal/mol). For MULL(solv), the blue squares, all results lie de facto within the ±0.25 kcal/mol threshold ($\delta\Delta A$(**5-t2**) $= -0.28$ kcal/mol, $\delta\Delta A$(**Ser**) $= -0.26$ kcal/mol). It should be noted that taking the statistical uncertainty of the results into account, none of these deviations from the ±0.25 kcal/mol threshold are statistically significant. Finally, all MULL(solv*) results (green diamonds) lie well within the ±0.25 kcal/mol threshold.

Figure II.3 shows that in most cases, the use of a hybrid intermediate charge state lowers $\delta\Delta A$, and typically, the use of MULL(solv*) leads to the best results, followed in this order, using MULL(solv) and MULL(gas). To quantify this, we report in Table II.1 the mean absolute deviation (MAD) for the data shown in Figure II.3, together with the spread of $\delta\Delta A$ for each of the protocols. Already, the use of MULL(gas) lowers the MAD from 0.29 to 0.12 kcal/mol, and the largest error $\delta\Delta A$ reduces from 1.80 to 0.49 kcal/mol. The MULL(solv) intermediate state

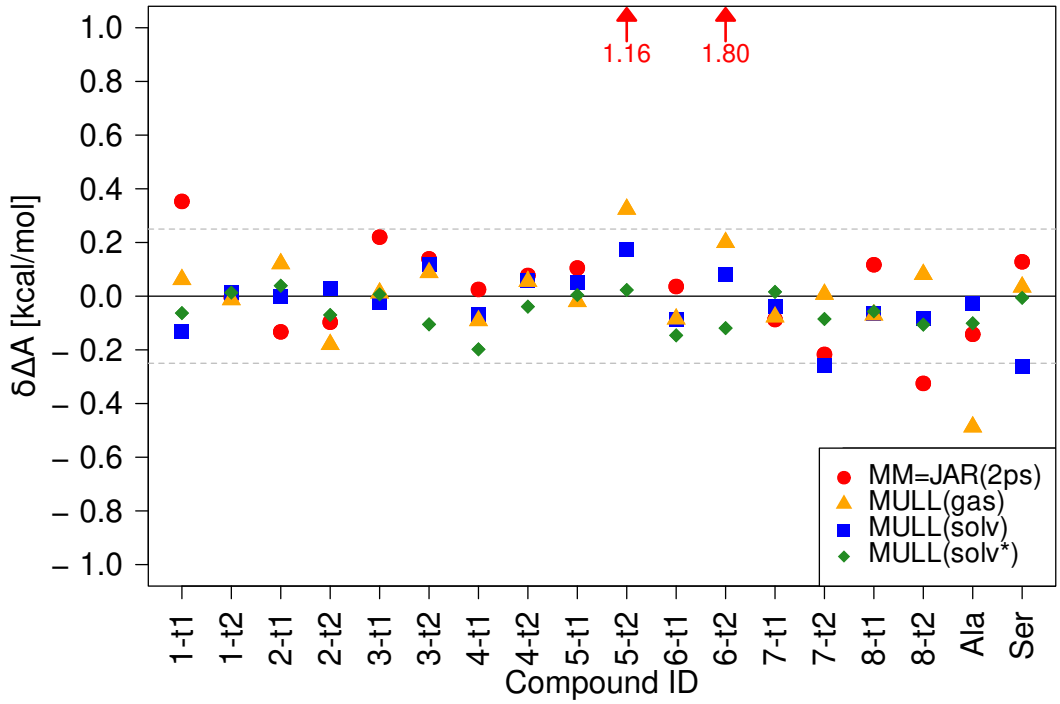


Figure II.3: Comparison of $\delta\Delta A$ obtained with three hybrid intermediate charge states: MULL(gas), orange triangles; MULL(solv), blue squares; MULL(solv*), green diamonds. All results were obtained from 200 NEW switches of 2 ps length to compute $\Delta A^{\mathrm{MULL}\rightarrow\mathrm{SQM}}$ and include the correction $\Delta A^{\mathrm{MM}\leftrightarrow\mathrm{MULL}}$; cf. Equation (II.2b). The JAR(2ps) results already shown in Figure II.2 (red circles) are included for comparison purposes.

leads to a further improvement (MAD = 0.09, and the largest error is 0.28 kcal/mol). Finally, the MAD for MULL(solv*) is 0.07 kcal/mol and the largest $\delta\Delta A$ is as low as 0.20 kcal/mol.

The improvements resulting from the hybrid intermediate charge states are roughly inversely proportional to the computational effort to obtain the average Mulliken-like charges. The MULL(gas) and MULL(solv) charges are obtained from configurations at the MM level of theory; the required computations at the SQM(/MM) level of theory to obtain the Mulliken charges are slightly more costly for the solvated system. By contrast, to obtain the MULL(solv*) charges, sampling needs to be carried out at the high (SQM/MM) level of theory. Given that the improvements of using MULL(solv*) rather than MULL(solv) are relatively small, this leads to an additional computational cost that, at least for the systems studied here, seems not to be worthwhile. In other words, the use of the MULL(solv) hybrid partial charges seems to be a good compromise between correctness and computational cost. Each of the indirect NEW switching protocols requires the calculation of

Table II.1: MAD (mean absolute deviation) for the results shown in Figure II.3, as well as the spread of $\delta\Delta A$ (respectively, the smallest and largest absolute deviations).

Pathway	**MAD [kcal/mol]**	**Spread MAD [kcal/mol]**	
		Min	**Max**
MM	0.29	0.01	1.80
MULL(gas)	0.12	0.01	0.49
MULL(solv)	0.09	0.01	0.28
MULL(solv*)	0.07	0.01	0.20

$\Delta A^{\mathrm{MM}\leftrightarrow\mathrm{MULL}}$; however, this is a strict force field-based free energy calculation, which can be computed quickly and efficiently on GPUs using the CHARMM/OpenMM interface (cf. Section II.4.4).

II.2.4 Performances of Modified Switching Protocols

Figure II.4 summarizes the results obtained with stepwise linear switching protocols for a subset of the model compounds. For details, see Section II.4.3 and Figure 11. In all cases, the baseline results, indicated as red, solid circles, are the free energy difference obtained from one-step MM $\rightarrow$ SQM switches, i.e., the MM/JAR(2ps) results already presented. For the first problematic case, **5-t2**, all three stepwise linear protocols reduce $\delta\Delta A$ considerably. However, in the case of **6-t2**, it is difficult to speak of improvements. Two protocols, **L2-1** and **L3-1**, lower $\delta\Delta A$, but the values are still far outside of the $\pm$0.25 kcal/mol threshold. The third protocol, **L3-2**, which worked extremely well for **5-t2**, actually increases $\delta\Delta A$ for **6-t2** compared to **L1**, the default linear protocol. The other three compounds included in the subset were chosen as negative controls because they already performed well with the JAR(2ps) protocol. The performance of the stepwise linear protocol is comparable, except for one poor result for **4-t2** when using **L3-1**.

Turning to the combination of stepwise linear and indirect NEW switching protocols with the MULL(solv) hybrid intermediate charge state (shown as squares), most of the results lie within the $\pm$0.25 kcal/mol threshold. However, since in this case, the performance of the linear protocol was excellent to begin with, any improvements are marginal. Moreover, for **4-t2**, the stepwise linear protocols perform slightly poorer, and for **6-t2**, $\delta\Delta A$ obtained with the **L3-1** is ¿1 kcal/mol.

Based on these findings, the performances of the stepwise linear switching protocols seems mixed at best. None of the protocols tested consistently improved

the results. Furthermore, even when the results were improved and fell within the ±0.25 kcal/mol threshold, the standard deviation of the work values, σ_W, was always larger than for the linear switching protocol. Since σ_W is a sensitive indicator of whether one can expect the convergence of JAR calculations [32], the utility of the protocols that increase it seems doubtful.

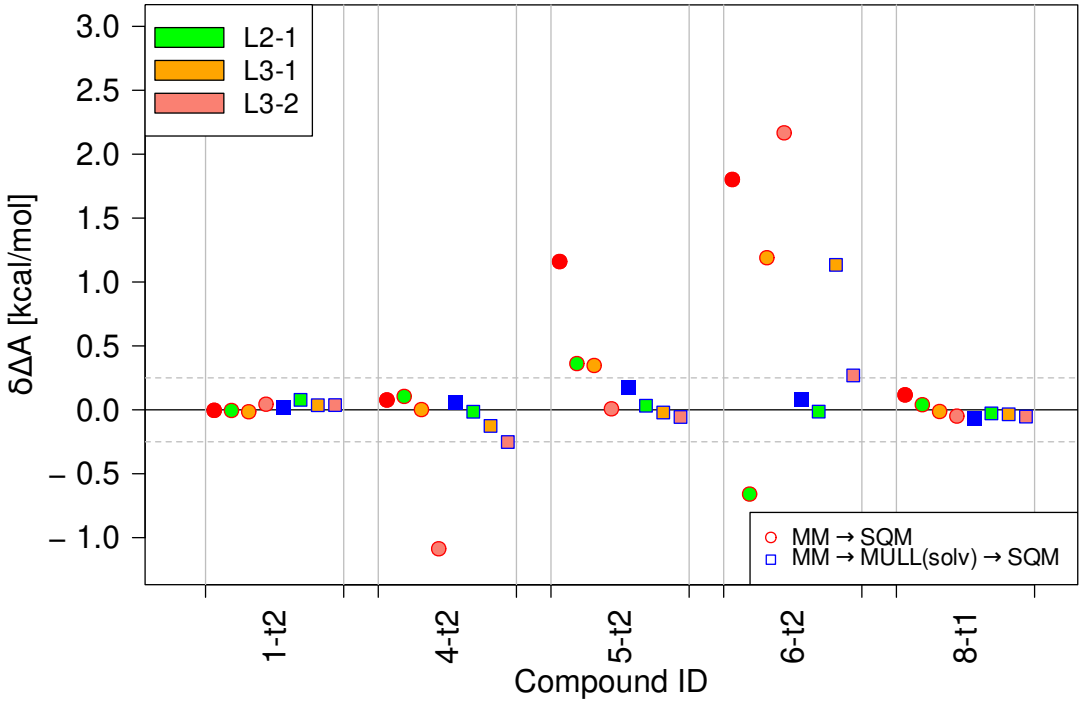


Figure II.4: Performances of the three modified switching protocols for a subset of the compounds. Results for the direct NEW switching protocols MM → SQM are shown as red circles; the results for calculations using the MULL(solv) hybrid charge intermediate state are blue squares. The fill color of the circles/squares indicates the stepwise linear switching protocol used (green = **L2-1**, orange = **L3-1** and salmon = **L3-2**). For comparison purposes, the MM/JAR(2ps) (red circles) and MULL(solv) results (blue squares), both using the default switching protocol **L1**, are included as well.

II.2.5 A Detailed Analysis of the Factors Affecting Convergence

Effects of Charge Distribution on Solute Properties

To quantify the differences between the various charge models used in this study, we calculated the root-mean-square deviation of the atomic charge differences $\mathrm{RMSD_q}$ between the MM, MULL(gas), MULL(solv) representations, using the MULL(solv*)

charges as the reference value (see Section II.4.3, Equation (II.9)). Since the latter are the average of the Mulliken charges derived from simulations at the SQM/MM level of theory, they are closest to the fluctuating charge distribution at the target high level of theory. The variability of the Mulliken charges derived from the SQM(/MM) reevaluations about their mean values were always extremely low, fluctuating typically less than ±0.03 e about the mean; the largest value observed was 0.08 e. These low standard deviations make the use of average values meaningful in the first place. In addition to $\mathrm{RMSD_q}$, we looked at the magnitude of the differential dipole moment $|\Delta\vec{\mu}| = \Delta\mu$ (Equation (II.10)) and the angle Θ_{SQM} between the dipole moment of a charge distribution and that of the MULL(solv*) reference charges (Equation (II.11)).

The individual results for each compound are listed in Table B.12 and depicted in Figures B.4, B.5 and B.6 of B. Some overall trends can be seen from the MAD values of $\mathrm{RMSD_q}$, $\Delta\mu$, and Θ_{SQM} in Table II.2. As expected, the MAD($\mathrm{RMSD_q}$) is largest for MM and smallest for MULL(solv), with MULL(gas) in between; the same is the case for MAD($\Delta\mu$). This is in accord with the MAD($\delta\Delta A$) results of Table II.1 in Section II.2.3, where the use of the MULL(gas) intermediate charge state led to a noticeable improvement, and the MULL(solv) charges lowered the MAD($\delta\Delta A$) even further. The MAD(Θ_{SQM}) behaves somewhat differently; here, the MULL(gas) value is lower than that of MULL(solv). As can be seen in Table B.12 and Figure B.6, this holds true for most compounds. Nevertheless, Θ_{SQM} of MULL(solv) is always smaller than that of the MM charges. The findings strongly suggest that the differences between the charge distribution at the two levels of theory have a strong influence on the convergence of the calculations of $\Delta A^{\mathrm{high}\rightarrow\mathrm{low}}$ via JAR, with the difference in magnitude of the two dipole moments being more important than the orientation of the respective dipole moment vectors.

Table II.2: MAD (Mean Absolute Deviation) over all compounds for $\mathrm{RMSD_q}$, the magnitude of the differential dipole moment $\Delta\mu$, and the angle Θ_{SQM} between the dipole moment vector of the respective charge distribution and that of the MULL(solv*) charge distribution.

Pathway	**MAD $\mathrm{RMSD}_{\boldsymbol{q}}[\boldsymbol{e}]$**	**MAD $\boldsymbol{\Delta\mu}$ [D]**	**MAD $\boldsymbol{\theta_{SQM}}[^\circ]$**
MM	0.14	3.38	29.8
MULL(gas)	0.06	2.44	7.6
MULL(solv)	0.04	1.81	16.0

Based on the overall convergence results (Figure II.2), we surmised that compounds containing a lactam moiety tended to converge more slowly than the cor-

responding lactim state, e.g., **5-t2** vs. **5-t1**, etc. One reason for this is that the differences between the MM and the SQM(/MM) descriptions are greater for the lactams than for the lactims. The analysis of partial charges points in this direction as well. Several differences can be discerned in the detailed charge distribution data for the individual molecules, e.g., Figure B.4. For all lactim–lactam pairs (compounds **2**, **3**, **4**, **5**, **6**, **8**),the MM $\mathrm{RMSD_q}$ of the lactam state is noticeably higher than that of the corresponding lactim state. Similarly, the MM $\mathrm{RMSD_q}$ of each of the lactams (red circle) is at least 0.1e higher than for MULL(solv) (blue squares).

In Figure II.5, we visualize some points made above for the two molecules, for which obtaining a converged $\Delta A^{\mathrm{MM}\rightarrow\mathrm{SQM}}$ was most difficult, **5-t2** (top) and **6-t2** (bottom). For each of the three charge representations, MM, MULL(gas), and MULL(solv), we display the differential atomic charges, both as labels, as well as a color gradient from blue to red, and the resulting differential dipole moment vectors (orange arrows). The magnitude of the differential dipole moment $\Delta\mu = |\Delta\vec{\mu}|$ and the angle Θ_{SQM} between the dipole moment of a charge distribution and that of the MULL(solv*) reference charge distribution are listed directly. The difference in partial charges and the MULL(solv*) charges can be large, e.g., the difference for the -C=O part of the lactam moiety is almost ±0.5e for **5-t2** (top, left in Figure II.5). One further sees that the charge differences become smaller between the MM and the two MULL charge sets, with atoms colored in clear blue or red for MM; e.g., the -C=O group (left side) has just a shade of blue or red for MULL(solv) (right side). Accordingly, the length of the orange arrows, i.e., $\Delta\mu$ of the MULL charge states, is noticeably smaller than that for MM. Even more detailed plots, including the exact values of the charge difference for each atom, for **3-t2**, **4-t2**, **5-t2**, **6-t2**, and **8-t2**, can be found in Figures B.7–B.11.

Effects of Charge Distribution on the First Solvation Shell

Figure II.6 displays the difference between the average number of waters ΔN_{Waters} within ≤ 3 Å of the solute. The numbers are relative to the target high-level state, i.e., the average number of waters found in the SQM/MM simulations. First, one sees that ΔN_{Waters} for the MULL(solv*) charges (green diamonds), which were also derived from the SQM/MM simulations, are very close to zero, in accord with expectation. Several results for the MULL(solv) charges (blue squares) are also quite close to zero, but mostly for the lactams (particularly **2-t2**, **3-t2**, **4-t2**, **5-t2**) the difference in water molecules $\Delta N_{Waters} \approx -1$. For the MULL(gas) charges (orange triangles), all ΔN_{Waters} values are negative, i.e., on average, there are fewer water molecules in close contact with the solute, compared to SQM/MM. This is not too

surprising, because the charges were derived from SCC-DFTB gas phase calculations; hence, the solute did not experience polarization from its interaction with the solvent. This may also explain why on average, the use of the MULL(gas) hybrid intermediate state performed worse than MULL(solv), even though it performed significantly better than MM (cf. Table II.1).

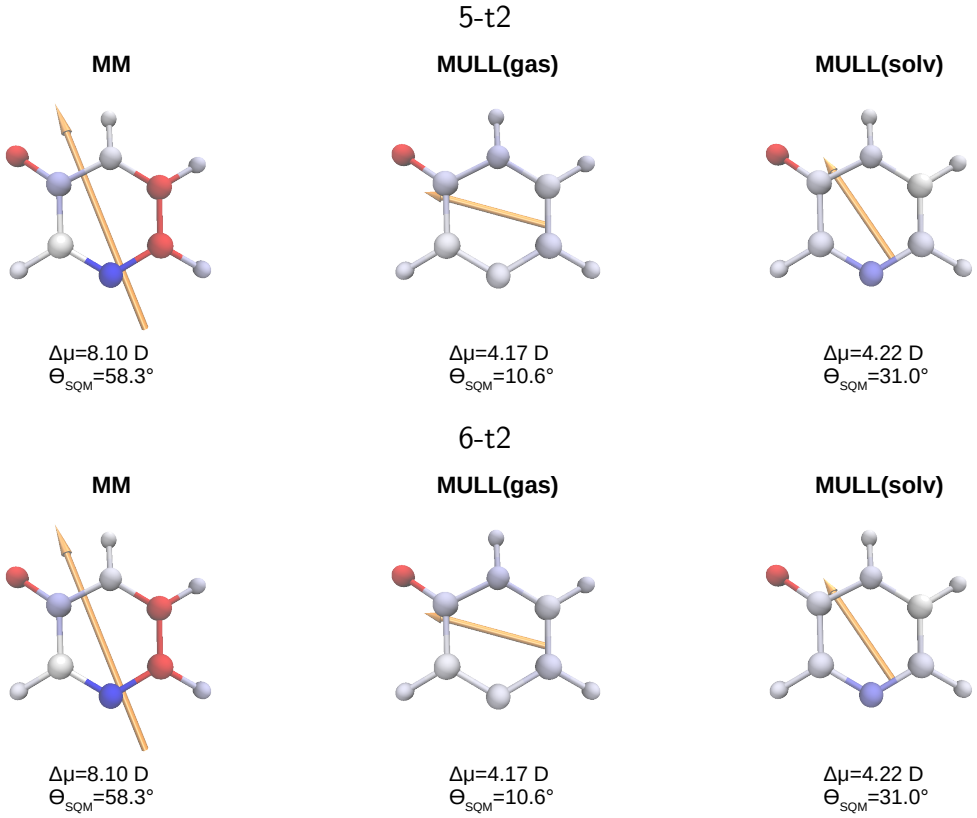


Figure II.5: Graphical representation of differences in partial atomic charges compared to MULL(solv*) for **5-t2** (**top**) and **6-t2** (**bottom**). Charge differences are indicated as a color gradient from blue ($\delta q = -0.2e$) to red ($\delta q = +0.4e$). The differential dipole moment $\Delta\vec{\mu}$ is displayed as an orange arrow. The magnitude $\Delta\mu = |\Delta\vec{\mu}|$ and the angle Θ_{SQM} between the dipole moment of the respective charge distribution with that of the MULL(solv*) reference distribution are given below each structure.

The MM results (red circles) fall into two distinct groups, with either more or less water present than in SQM/MM. For most lactams, with the slightly surprising exception of **6-t2**, ΔN_{Waters} is negative, whereas the lactims, i.e., have a positive ΔN_{Waters}. Thus, one sees again distinct differences in properties for the force field representations of lactims and lactams, respectively.

Water Reorientation Dynamics

In the previous subsections, we characterized the influences of different charge distributions (atomic partial charges) on the static properties of the solute and the

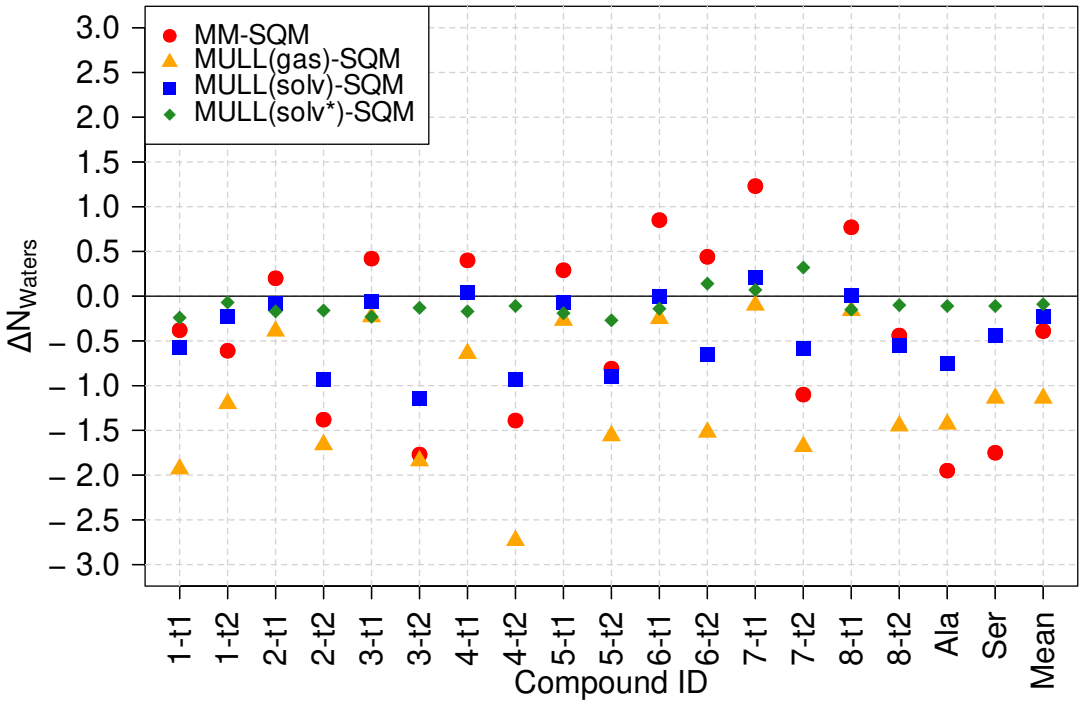


Figure II.6: Difference in the average number of water molecules closer than 3 Å, compared to SQM/MM. The average number of water molecules in the SQM/MM simulations was used as the reference value.

surrounding solvent layer in several ways. As the endpoints of the NEW switches have different charge distributions, it is of considerable interest to study the dynamics of solvent reorientation. We therefore computed the unnormalized Stoke shift relaxation $\overline{\Delta\Delta U(t)}$ for several compounds, and the MM and MULL(solv) charge distributions (see Section II.4.3, Equation (II.13)). The results for **5t-2** are shown in Figure II.7; analogous plots for **4t-2**, **6t-2**, and **8t-2** can be found in Figures B.12 B.13 and B.14 in B. All fit parameters are summarized in Table B.13 of B.

As described in Section II.4.3, the mono-exponential fit was carried out in such a way that it picks out the slow process(es) of water reorientation. Looking at Figure II.7, one sees that the relaxation times for MM and MULL(solv) are relatively comparable ($\tau \approx 1$ ps), but that the prefactor ΔU_0 is quite different (¿10 kcal/mol for MM and ≈4 kcal/mol for MULL(solv)). Thus, while $\overline{\Delta\Delta U(t)}$ has for all practical reasons reached 0 after about 3 ps in the MULL(solv) case (blue curve), for MM (red curve), this is the case only after 5 ps. Since we would ideally carry out NEW switches of only 2 ps length, the obvious question is the value of $\Delta U(t = 2\,\text{ps})$. Using the fit parameters, we find that MM ≈ 1.7 kcal/mol, and that MULL(solv) ≈ 0.5 kcal/mol. These values roughly mirror the values for $\delta\Delta A$ for MM/JAR(2ps)

and MULL(solv), respectively.

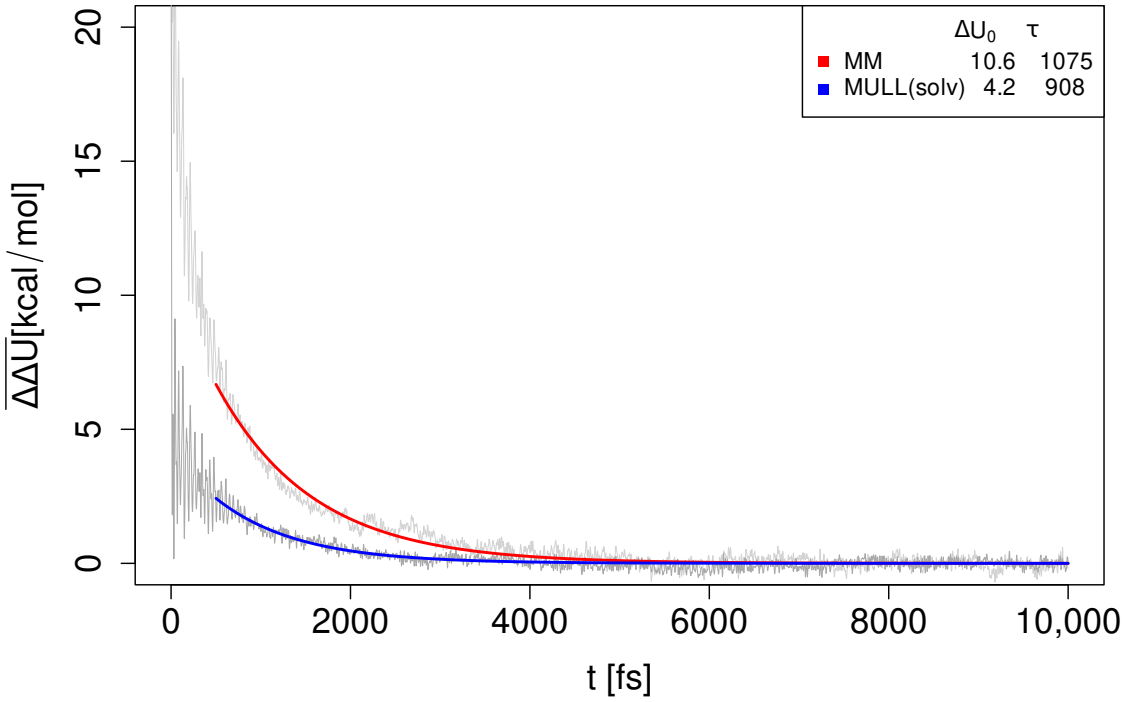


Figure II.7: Water reorientation dynamics of TIP3P solvent for **5-t2** when switching from MM and MULL(solv) to SQM/MM. Raw data are shown in gray; fit for MM in red and fit for MULL(solv) in blue. The fit parameters (ΔU_0 in kcal/mol, τ in fs) are listed in the inset; cf. Equation (II.14).

For the second problem case, **6-t2**, very similar results were obtained; see Figure B.13 and Table B.13.

Overall, the slow decay process of $\overline{\Delta\Delta U(t)}$ always occurs with a time constant $\tau \approx 1$ ps. Thus, the degree to which the system is still out of equilibrium regarding the SQM/MM target state therefore depends crucially on how big the difference is at $t = 0$. Thus, it becomes clear that convergence is facilitated if the solute charge distributions of the initial and final state resemble each other, explaining why better results were obtained for all the MULL intermediate charge states. Phrased differently, to obtain converged JAR results from 2 ps switching simulations, the charge distributions of the initial and final states have to be very similar. Since the determining factor seems to be the initial difference in charge distribution, using stepwise linear switching paths cannot help much, explaining at least in part why we did not obtain any real improvements from the stepwise linear protocols we attempted.

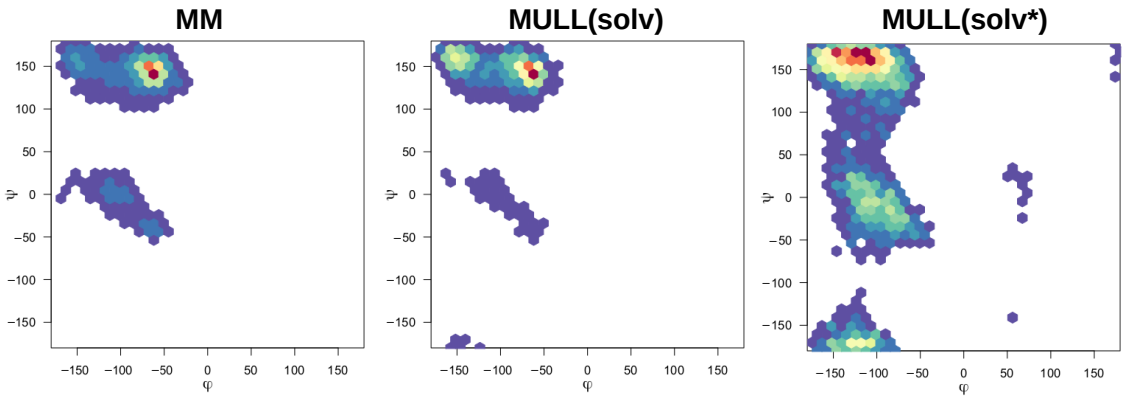


Figure II.8: Ramachandran plots of pseudo-dipeptide/blocked alanine.

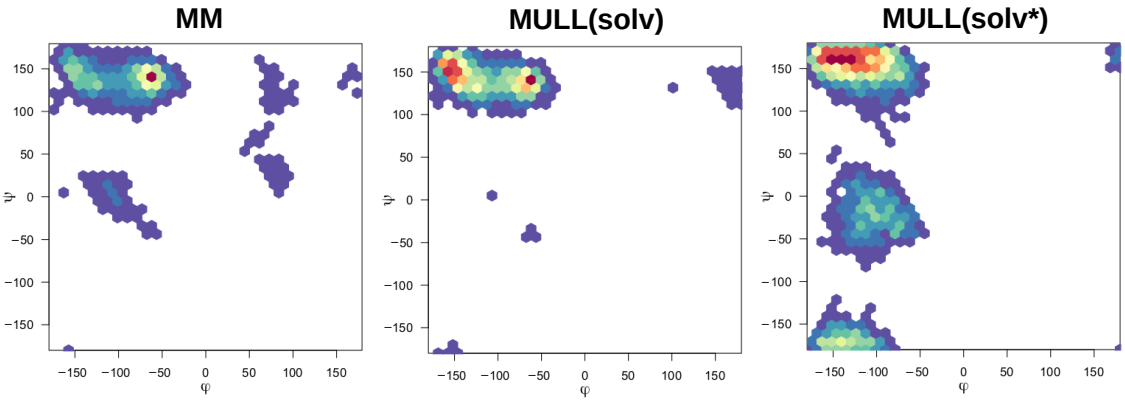


Figure II.9: Ramachandran plots of pseudo-dipeptide/blocked serine.

Interplay between Charge Distribution and Conformational Preferences
The compounds used in this work were mostly rigid and specifically chosen to avoid complications from conformational degrees of freedom, the exceptions being **1-t1** and the blocked amino acids **Ala** and **Ser**. Modifying partial charges as for the MULL hybrid intermediate states may have an effect on conformational preferences. As can be seen in Figure II.3, all three MULL hybrid intermediate states improve the convergence for **1-t1**. For the two blocked amino acids, MULL(gas) results in a poorer convergence for **Ala**, and MULL(solv) results in a slightly poorer convergence for **Ser**. Given that **Ala** and **Ser** are the smallest possible peptide-like molecules with protein backbone-like ϕ and ψ dihedral angles, we performed some analyses on the dihedral angle conformational preferences. In earlier work [31], we showed that purely classical Hamiltonians and SQM/MM Hamiltonians resulted in different preferred conformations of ϕ and ψ, as well as χ_1 in the **Ser** case. In Figures II.8 and II.9, ϕ/ψ maps for **Ala** and **Ser** are shown for MM, MULL(solv), and SQM. The differences between MM (left) and SQM (right) are clearly visible. For both blocked amino acids, MM has a single narrow minimum at $\phi \approx 150°$, $\psi \approx -50°$, whereas SQM has

a broader distribution at $\phi \approx 150°$, and $-150° < \psi < -50°$. While the MULL ϕ/ψ maps (middle plots) are more similar to MM than to SQM, a second minimum at $\phi \approx 150°/\psi \approx -150°$ has appeared. Thus, although the effect is small, the use of hybrid charge intermediates also makes this state slightly more similar to a high-level state, in terms of conformational preferences.

II.3 Discussion

The optimized protocol proposed in Ref. [36] (200 switches of 2 ps length) would be efficient enough for many practical applications. However, it has only been validated in the gas phase. If water reorientation is responsible for the slower convergence in aqueous solution, as observed previously [31, 37], then, given the experimentally known slower relaxation time constant of >1 ps, switching lengths of 2 ps may not be long enough. This suspicion was confirmed in this work, and we found that in one case, even a switching length of 5 ps was insufficient.

We explored two approaches to keep the switching length at 2 ps : (i) introducing a hybrid intermediate charge state, which makes the solute charge distribution more similar to that at the SQM/MM level of theory. (ii) Since the solvent reorientation also has an ultrafast component, we used a stepwise linear switching scheme instead of the standard one, switching faster, initially followed by one or two slower steps. While the use of hybrid intermediate charge states significantly improved convergence, the benefits of multi-stage switching protocols were mixed.

The partial charges of the intermediate hybrid states were obtained by averaging over the Mulliken charges obtained from 200 energy evaluations at the SQM level of theory. The computational cost of this is relatively small; the overhead, if any, comes from sampling the configurations over which one averages in the first place. Not surprisingly, the best results (lowest overall $\delta\Delta A$; see Table II.1) were obtained from coordinate sets generated during SQM/MM simulations, i.e., calculations at the target high level of theory (MULL(solv*)). While we have performed such calculations to obtain reference results with CRO, we would ideally prefer to avoid them in practical applications. In contrast, the cheapest approach is to use the coordinate sets saved during the MM gas phase simulations (MULL(gas)). While these results were an improvement over the direct MM→SQM/MM switches, the MAD($\delta\Delta A$) was higher than that of MULL(solv*). The results for the third method tested, MULL(solv), fell in between. Here, the configurations to be reevaluated at the SQM/MM level of theory were generated from the MM simulations in the solvated phase. Since one can reuse the stored coordinates/restart files to start

the NEW switching simulations, the computational cost is practically zero. All results obtained with the MULL(solv) hybrid intermediate charge state met our rather stringent quality criteria; the improvements over the direct MM→SQM/MM switches were noticeable, and the difference in convergence over the more expensive MULL(solv*) approach was small (a MAD of 0.09 instead of 0.07 kcal/mol). Thus, the MULL(solv) approach to derive an intermediate charge representation seems to be a good compromise between computational cost and convergence improvement. While for this study, averaged Mulliken charges were the logical choice because of the use of SCC-DFTB, other charge assignment procedures could be explored if different QM methods were used, such as ESP or CM5-symmetrized charges [43, 19]. All methods require an additional, free energy simulation to obtain $\Delta A^{\mathrm{MM}\leftrightarrow\mathrm{MULL}}$ at the low level of theory, but its computational cost is small compared to that of the NEW switching simulations needing SQM/MM Hamiltonians.

As reported in Section II.2.4, we have not seen consistent convergence improvements for all systems using the two-stage and three-stage switching protocols. This prompted a detailed analysis of how the MM and SQM/MM representations differ. Comparing the MM charges of the force field to the average Mulliken charges, one sees that there are indeed surprisingly large differences in charge distribution, and hence, the solute dipole moment. The differences in properties of the solute have a noticeable effect on the number of water molecules in close contact with the solute. Furthermore, by studying the time correlation function $\overline{\Delta\Delta U(t)}$ after switching from one level of theory to the other, we were able to confirm the time scales of experimental solvation spectroscopy. There clearly is a fast process with a relaxation time $\ll 1$ ps, and a slower process with $\tau \approx 1$ ps. This slower time constant does not change much if the initial charge distributions are equal to MM or MULL(solv). If the difference between the initial and final distribution is large, then the effect of the unfinished solvent reorientation after 2 ps (the switching length we are aiming for), will also be large. If the two charge distributions are more similar, as is the case for MULL(solv), then $\Delta U(t = 2 \text{ ps})$ is small enough to have little effect on the results, even though the solvent reorientation may not be complete. On the one hand, this observation explains the effectiveness of using hybrid charge intermediate states. On the other hand, it helps to rationalize our failure to obtain consistently convergent results with the multistage switching schemes. While one could certainly construct more complex switching paths, the speed of the solvent reorientation would remain the same.

II.4 Materials and Methods

II.4.1 Computing Free Energy Differences between the Levels of Theory Using NEW Methods

In several previous studies, [30, 31, 32, 33, 34, 35, 36] we demonstrated that NEW methods are a robust means of computing free energy differences between levels of theory, such as an MM and an SQM/MM description of a system. Given a series of non-equilibrium work $W^{MM\to SQM/MM}$ values obtained from switching simulations, the free energy difference is computed according to Jarzynski's equation [40]:

$$\Delta A^{MM\to SQM/MM} = -k_B T \ln \left\langle \exp(-\beta W^{MM\to SQM/MM}) \right\rangle \tag{II.3}$$

where k_B, T, and β have the usual meanings of Boltzmann's constant, absolute temperature, and $\beta = 1/k_B T$. The angular brackets indicate averaging over a sufficiently large number of individual work values. To realize this scheme, one has to save snapshots during an equilibrium MD simulation at the MM level of theory. These serve as the starting point for the switching simulations using a hybrid energy function:

$$U(\lambda_t) = (1-\lambda_t)U^{MM} + \lambda_t U^{SQM/MM} \tag{II.4}$$

Here, $0 \leq \lambda_t \leq 1$ is the coupling parameter which is incremented, e.g., linearly at each step of the switching simulation, according to

$$\lambda_t = \lambda_t(i) = i/N_{switch}, \qquad i = 1, 2, \ldots, N_{switch} \tag{II.5}$$

In our protocols, N_{switch} is, e.g., 2000 or 5000 steps, corresponding to a switching length of 2 or 5 ps with a timestep $\delta t = 1$ fs, and we carry out two hundred of such switches (see below for details). Equation (II.5) describes the simplest switching protocol; in this work, we also test modified protocols where λ_t is incremented at different speeds. The switching simulations are costly because the evaluation of the mixed potential energy function $U(\lambda_t)$ requires the high-level energy function $U^{SQM/MM}$ and its gradient. However, since the switches are independent, they can be carried out in parallel. For each timestep at which λ_t is incremented, the work is accumulated according to (cf. Ref. [44])

$$W(t+\delta t) = W(t) + U(x_{t+\delta t}, \lambda_{t+\delta t}) - U(x_{t+\delta t}, \lambda_t) \tag{II.6}$$

One-sided methods, such as Jarzynski's equation, are less efficient and reliable than

the two-sided approaches [45]. The process outlined above to obtain $W^{MM\to SQM/MM}$ can be reversed, so one ends up with distributions $P^{MM\to SQM/MM}(W)$ of work values in the forward, and $P^{SQM/MM\to MM}(W)$ in the backward direction. Crooks showed that these two distributions are related according to[41]

$$P^{MM\to SQM/MM}(W) = P^{SQM/MM\to MM}(-W)\exp[\beta(W - \Delta A^{MM\to SQM/MM})] \quad \text{(II.7)}$$

from which one can deduce the free energy difference between the two states, i.e., between the two levels of theory. The notation in Equation (II.7) has been adapted to the present application. The practical use of Crooks' equation is computationally costly, since one needs to generate equilibrium configurations at the high level of theory. As in the past, Crooks' equation will therefore only be used to generate reference results to assess the convergence of free energy differences obtained using Jarzynski's equation. In this work, Jarzynski's and Crooks' equations (or their use) are abbreviated as JAR and CRO, respectively.

The mixing of Hamiltonians as required by Equation (II.4) was realized with the MSCALE module of CHARMM [46], which makes it possible to combine energies and forces from different sources, including separate programs, in an almost arbitrary manner. The MSCALE functionality is coupled to one of CHARMM's tools (PERT) to compute free energy differences [47]. Realizing that short/fast slow-growth thermodynamic integration calculations yield a non-equilibrium work rather than a converged free energy differences [48], one can exploit PERT's slow-growth mode to carry out NEW switches. Using PERT, it is also possible to carry out stepwise linear switching protocols; see Section II.4.3 below.

II.4.2 Choice of Model Systems

This study focuses on the optimization of the calculation of $\Delta A^{MM\to SQM/MM}_{X\ solv}$ (red arrow in Figure II.1a); in other words, the free energy difference between a system in which all interactions are described using classical mechanics, and a hybrid system, in which a small region of interest is described using SQM, and the remainder described using MM. Examples are solute–solvent systems, in which the solute is modeled either via molecular mechanics or via quantum chemistry, whereas the solvent (water) is always treated classically. As outlined earlier, difficulties may arise from the differences in the charge distributions of the MM and SQM representations of the solute molecule X. To exclude other factors hindering convergence, suitable model systems should be relatively rigid and not have multiple conformational substates.

Many tautomer pairs fulfil this requirement. Therefore, we picked seven such pairs from the tautomer part of the SAMPL2 challenge [49]. One additional pair was taken from *Tautobase* [50]. We also included N-acetyl-alanine-methylamide (**Ala**) and N-acetyl-serine-methylamide (**Ser**) as model solutes, since these were the two compounds for which we noticed slow convergences in Ref. [31]. All model compounds used in this work are shown in Figure II.10.

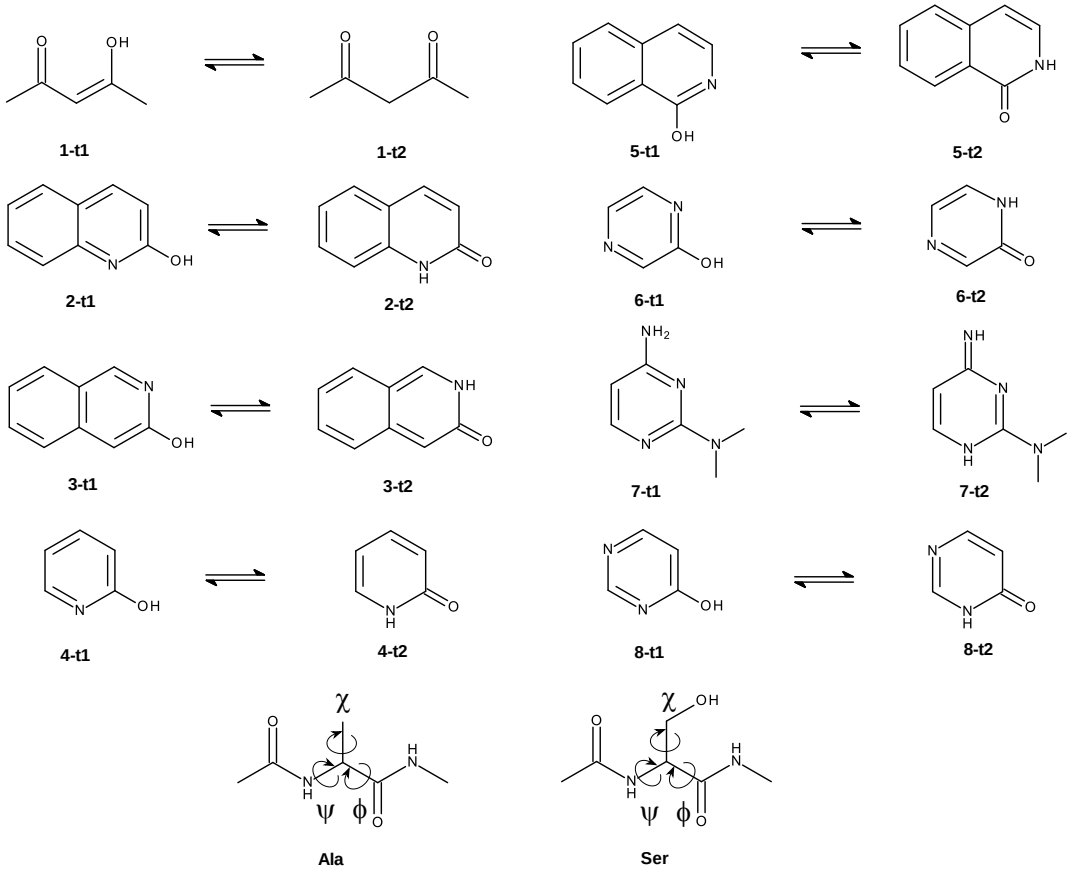


Figure II.10: Model systems used in this work: eight tautomer pairs, seven taken from the SAMPL2 challenge [49], and one (**7-t1** $\rightleftharpoons$ **7-t2**) from *Tautobase* [50]. In addition, we used the blocked amino acids **Ala** and **Ser** as model solutes.

The prediction of tautomer preferences in aqueous solution remains a challenging problem [51] and is out of the scope of this study. Here, we concentrate on finding efficient protocols to compute the correction $\Delta A_{X\ solv}^{MM \rightarrow SQM/MM}$, the red arrow in Figure II.1. Since (S)QM/MM methods are likely necessary for answering questions involving tautomer preferences, the present work establishes the necessary methodology to enable such calculations efficiently.

II.4.3 Strategies to Improve the Convergences of NEW Simulations

Hybrid Charge Intermediates

As outlined in the Introduction, cf. Figure II.1b, we investigated whether the quantity of interest $\Delta A_{X\ solv}^{MM \rightarrow SQM/MM}$ can be computed more efficiently by inserting an intermediate state. We refer to it as "*MULL*"; the rationale for this label will become clear shortly. Thus, we have

$$\Delta A_{X\ solv}^{MM \rightarrow SQM/MM} = \Delta A_{X\ solv}^{MM \leftrightarrow MULL} + \Delta A_{X\ solv}^{MULL \rightarrow SQM/MM} \tag{II.8}$$

This intermediate state should have a charge distribution resembling that of the SQM/MM end state. To be useful in practice, the free energy difference between the initial MM and the intermediate state $\Delta A_{X\ solv}^{MM \leftrightarrow MULL}$ must be easy (=fast) to compute. In practice, this means that the interactions at the intermediate state must also be classical. The obvious choice, therefore, is to use the force field of the initial MM state with suitably modified partial charges.

In all our methodological work to date [30, 31, 32, 33, 34, 35, 36], we used the SCC-DFTB SQM method [52, 53] as the high level of theory.

On the one hand, SCC-DFTB is expected to be sufficiently similar to ab initio QM methods, so that insights obtained will be transferable to full DFT QM methods. On the other hand, SCC-DFTB is fast enough to explicitly carry out simulations at the high level of theory, something which is not feasible for ab initio QM methods. The ability to carry out simulations at the SQM/MM level of theory makes it possible to obtain rigorous reference results using Crooks' equation; cf. the previous section.

Since Mulliken charges and Mulliken charge analysis are at the core of the SCC-DFTB methodology [52, 53], a logical choice for the intermediate state is the combination of the MM force field (see below for details) with fixed partial charges derived from the Mulliken charges used at the SCC-DFTB level of theory; hence, the superscript MULL in Figure II.1b and Equation (II.8). In SCC-DFTB calculations, the converged Mulliken charges change at every simulation step; therefore, an averaging procedure is needed. There are several possibilities for how to obtain a representative sample of configurations for which Mulliken charges are computed. The following three approaches are tested in this study. (i) Clearly, the most exact method is to use simulations at the high level of theory (the solute treated using SCC-DFTB, surrounded by MM waters) to obtain configurations for which the Mulliken charges are calculated and then averaged. All simulations and results using this intermediate state are labeled *MULL(solv*)*. Our goal, however, is to avoid

simulations at the high level of theory as much as possible. Therefore, we also used configurations sampled at the MM level of theory to derive Mulliken charges for the intermediate representation. Specifically, we used (ii) the Mulliken charges obtained from snapshots of the MM gas phase simulation of the solute (labeled as *MULL(gas)*), or (iii) we used snapshots from the corresponding MM simulations in water to obtain Mulliken charges for the solute (labeled as *MULL(solv)*).

The three sets of MULL charges were obtained by averaging over configurations saved during equilibrium MD simulations, i.e., MULL(gas) charges were derived from the MM gas phase simulations, MULL(solv) from the MM simulations in solutions, and MULL(solv*) from the SQM/MM simulations in solution. These are the simulations needed to save restart files for the NEW switching simulations to the respective other levels of theory. As described below (cf. Section II.4.4), we always carried out eight repetitions, i.e., eight MD simulations per state and system. From each of these, 25 configurations were taken and reevaluated at the SQM(/MM) level of theory, writing out the converged Mulliken charges. Thus, each MULL charge set was obtained by averaging over 200 configurations.

If one inserts one of the MULL intermediate states to compute $\Delta A^{MM \rightarrow SQM/MM}$ as described above, then one also must compute $\Delta A_{Xsolv}^{MM \leftrightarrow MULL}$ to close the thermodynamic cycle (cf. Figure II.1b). Since the MM and the MULL states differ only in the partial atomic charges, we calculated $\Delta A_{Xsolv}^{MM \leftrightarrow MULL}$ using equilibrium free energy methods. Specifically, in addition to the end states, we used three equidistant intermediate states and calculated the free energy difference using Bennett's acceptance ratio method (BAR) [42].

Stepwise Linear Switching Protocols

All protocols that are employed to switch from MM to SQM/MM are quite rapid, so the system will not remain in equilibrium. In the context of slow-growth thermodynamic integration, this has been referred to as Hamiltonian lag, and has been shown to lead to poorly converged results [54, 55, 48]. Although JAR [40] makes it possible to obtain the equilibrium free energy difference for a process from multiple irreversible work values, the convergence of the results depends on the variance of the work values, which in turn depends on the deviation of the switching trajectories from the equilibrium conditions [56, 44].

In addition to inserting an intermediate state with partial charges resembling the high level of theory, as just described, we explored the use of modified switching protocols. Results from solvation spectroscopy [38, 39] indicate that the faster process of water reorientation in response to a change in the charge distribution of a

solute is completed after approximately 0.5 ps. We therefore tested piecewise linear switching protocols consisting of two or three stages, referred to as L2-X and L3-X, respectively. The three protocols investigated in detail are shown in Figure II.11, which also include the default linear protocol L1 of Equation (II.5) (leftmost plot). In all cases, the total switching length τ is 2000 fs. In L2-1, instead of incrementing λ_t linearly, there are two stages: first, λ_t is incremented from $\lambda_t = 0.00 \rightarrow 0.35$ in 200 fs, followed by slower linear switching from $\lambda_t = 0.35 \rightarrow 1.00$ over 1800 fs. The two L3-X protocols use three stages. In L3-1, λ_t is switched from 0.0 to 0.35 to 0.8 to 1.0 over 200, 900, and 900 fs, respectively. Finally, L3-2 is very similar, but λ_t is switched from 0.0 to 0.35 to 0.8 to 1.0 over 200, 400, and 1,400 fs, respectively. In all modified protocols, λ_τ is initially changed more rapidly compared to L1, followed by a slower change in λ_τ during the remainder of the switch.

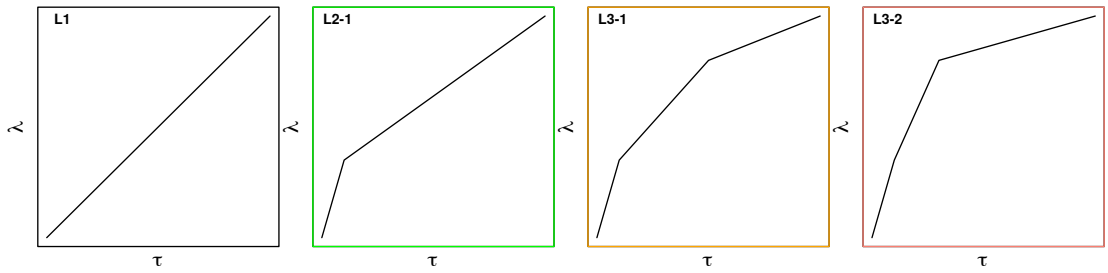


Figure II.11: The standard linear (L1) and three modified switching protocols investigated. The total switching length τ was 2 ps in all four cases. The color code used for L2-1, L3-1, and L3-2 is also used in Results.

To verify the applicability and the performance of the modified compared to the standard linear switching protocol, systematic tests were carried out for five compounds (**1-t2**, **4-t2**, **5-t2**, **6-t2**, and **8-t1**). This selection was motivated by the results obtained with the linear protocol, and includes both "easy" and "difficult" solutes with respect to convergence (cf. Results, Section II.2.2).

Analyses Carried Out

The main focus of this study lies on computing $\Delta A_{Xsolv}^{MM \rightarrow SQM/MM}$ efficiently, and most of our results are concerned with the accuracy of JAR-based protocols compared to the reference results obtained with CRO. Additional analyses and characterizations were carried out to understand how solute–solvent interactions change and thus affect convergence when switching from an MM to an SQM representation of the solute.

Characterizing Charge Distributions As described in Section II.4.3, we use three intermediate charge distributions to facilitate the transformation from the

low to the high level of theory. It is therefore of interest to quantify the differences between the charge sets used. In the following, the reference charge set is always the one that is derived from averaging over the Mulliken charges obtained from SQM/MM trajectories [MULL(solv*)]. To quantify the difference between two charge sets, we computed the root-mean-square deviation $\mathrm{RMSD_q}$, defined as

$$\mathrm{RMSD_q} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(q_i^{method} - q_i^{MULL(solv*)}\right)^2} \tag{II.9}$$

Here, N is the number of atoms of the molecule, the $q_i^{MULL(solv*)}$ charges serve as the reference, and the q_i^{method} is one of the other charge sets used; i.e., *method* can be MM, MULL(gas), or MULL(solv).

Because all our solutes are neutral, the most important quantity affected by a change in partial charges is the dipole moment $\vec{\mu}$. To compare the dipole moments resulting from the various charge sets, we consider the "differential" dipole moments between two charge distributions of a solute, i.e.,

$$\Delta\vec{\mu} = \sum_{i=1}^{N}(q_i^{method} - q_i^{MULL(solv*)}) \cdot \vec{r_i} \tag{II.10}$$

In practice, we mostly compare the length of this vector, i.e., $|\Delta\vec{\mu}| = \Delta\mu$.

Furthermore, the angle between the dipole moment vectors of two charge representations is of interest, which can be obtained in the usual way, is:

$$\Theta_{\mathrm{SQM}} = \cos^{-1}\left(\frac{\vec{\mu}^{method} \cdot \vec{\mu}^{MULL(solv*)}}{|\vec{\mu}^{method}| \cdot |\vec{\mu}^{MULL(solv*)}|}\right) \tag{II.11}$$

Dipole moments were computed using the `COOR DIPOle` command of CHARMM (see `https://academiccharmm.org/documentation/version/c47b1/corman` (accessed on 29 April 2023)); if the difference between two charge sets is assigned as partial charges, one obtains the corresponding differential dipole moment.

Characterization of the First Solvation Shell We analyzed all MM, MULL (gas), MULL (solv), MULL (solv*), and SQM/MM equilibrium trajectories to determine the average number of solvent water molecules in proximity to the solute. Specifically, a water molecule was considered to be in close contact if the distance between the water oxygen and a non-hydrogen atom of the solute was $\leq$ 3 Å. These analyses were carried out with the atom selection facility of CHARMM. (See `https://academiccharmm.org/documentation/version/c47b1/select` (accessed

on 29 April 2023)).

Dynamics of Solvent Reorientation To investigate the detailed dynamics of solvent reorientation upon changing the description of the solute from the MM to the SQM level of theory, we resort to ideas from computational spectroscopy, specifically the computation of the normalized Stokes shift $S(t)$ from non-equilibrium MD simulations. $S(t)$ is defined as [57, 58]:

$$S(t) = \frac{\nu(t) - \nu(\infty)}{\nu(0) - \nu(\infty)} \approx \frac{\overline{\Delta U(t)} - \overline{\Delta U(\infty)}}{\overline{\Delta U(0)} - \overline{\Delta U(\infty)}} \tag{II.12}$$

where $\nu(t)$ is the time-dependent frequency of the emitted fluorescence light of the probe molecule's chromophore. In computation, one assumes that changes in interactions between the solute and the solvent leading to $\nu(t)$ are purely electrostatic in nature. Thus, one takes equilibrated configurations from solute–solvent simulations, changes the partial charges from those describing the ground state to those describing the excited state, restarts the simulations, and monitors the difference in electrostatic solute–solvent interactions between the ground (U^0) and the excited state (U^*), $\Delta U(t) = U^*(t) - U^0(t)$ as a function of time. Just as the measured fluorescence signal is an average of the emissions from all the chromophores present in solution, one averages over at least a few hundred simulations; hence the bar in $\overline{\Delta U(t)}$.

In the context of our work, the low- and high-level representations of the solute (region of interest) play the roles of the ground and excited states in solvation spectroscopy. In both cases, the solvent water has to reorient itself following a change in the charge distribution of the solute. Therefore, we adopted the computational approach to compute $S(t)$ [57, 58] as follows. We used starting configurations equilibrated either at (i) the MM level of theory or (ii) the MULL(solv) intermediate state, and restarted simulations with the full SQM/MM description of interactions. For each system studied (**4-t2**, **5-t2**, **6-t2** and **8-t2**), we computed 800 simulations of 10 ps length; based on the experimental findings, it is reasonable to expect that solvent reorientation is completed after this time [38, 39]. At each timestep, we saved $\Delta U(t) = U^{SQM/MM}(t) - U^{method}(t)$, where *method* was either MM or MULL(solv). This task was facilitated by a locally modified version of CHARMM, but could equally well be accomplished through the post-processing of trajectories saved during the simulations. The 800 $\Delta U(t)$ time series were then averaged, resulting in the averaged time series $\overline{\Delta U(t)}$. Next, we estimated $\overline{\Delta U(\infty)}$, i.e., the limit of $t \to \infty$, by averaging (again) over the last 2000 entries of $\overline{\Delta U(t)}$ ($8\,\text{ps} \leq t \leq 10\,\text{ps}$).

Equation (II.12) is the definition of the normalized Stokes shift, whereas in the

present context omitting the normalization, it turned out to be advantageous. Thus, we operate with the averaged time series,

$$\overline{\Delta\Delta U(t)} = \overline{\Delta U(t)} - \overline{\Delta U(\infty)}, \tag{II.13}$$

which essentially is an unnormalized Stokes shift. To interpret it more easily, $\overline{\Delta\Delta U(t)}$ was fitted to a mono-exponential target function

$$\overline{\Delta\Delta U(t)} \propto \Delta U_0 \cdot e^{-t/\tau} \tag{II.14}$$

where ΔU_0 is the prefactor of the exponential decay at $t = 0$, t is the time in fs, and τ the relaxation time constant in fs. As described in the Introduction, it is known experimentally that the orientation of solvent water occurs on two timescales [38, 39]. Because any effects on the convergence of JAR calculations will result from the slower process, we intentionally always discarded the first 0.5 ps when carrying out the fit. Thus, the τ found using the fitting procedure should correspond to the slower reorientation process.

II.4.4 Overview of Simulations Carried Out

All simulations were carried out with CHARMM (developmental versions c44a2 and c47a1) [47]. The calculation of $\Delta A_{Xsolv}^{MM\rightarrow SQM/MM}$ requires equilibrium simulations to generate the configurations from which the NEW switching simulations are started. As described above, in addition to switching directly from the MM to the SQM/MM level of theory, we also inserted three force field-based intermediate states with different atomic partial charges. For each of these states, we also had to carry out equilibrium simulations, followed by NEW switches to the SQM/MM level of interest. The NEW switching simulations themselves were carried out either in a strictly linear fashion (protocol L1 in Figure II.11), or with a two- or three-staged linear protocol. To make sure that any convergence problems observed did not arise without solvent, we also carried out gas phase simulations for all compounds shown in Figure II.10 using the optimized protocol of Ref. [36].

Simulation Details

Preparation and Initial Equilibration For each molecule shown in Figure II.10, starting coordinates were generated using CHARMM-GUI [59, 60]. Missing force field parameters were generated using CGenFF 2.4.0 [61, 62, 63], as invoked by CHARMM-GUI. The solute–solvent systems were set up by placing each molecule

into a cubic box of water molecules with an initial side-length of 26 Å containing 572 (CHARMM-modified) TIP3 waters [64]. Any water molecule overlapping with a solute atom was deleted. Each system was then equilibrated as follows: 100 steps of steepest descent minimization were followed by a constant pressure/temperature (CPT) simulation of 100 ps length with a timestep of 1 fs, applying the Langevin piston barostat [65] (mass of the pressure piston: 400 Da, Langevin piston collision frequency: 20 ps^{-1}, Langevin piston bath temperature: 300 K). The final 20 ps of these equilibration runs were used to determine the average box size. In Table B.1 of B, we list the size of the simulation box determined in this manner, as well as the number of water molecules present, for each of the compounds studied.

Force Field-Based Equilibrium Simulations Starting from the initial solute coordinates and the equilibrated solute–solvent systems, eight Langevin dynamics (LD) simulations (timestep 1 fs, friction coefficient 5 ps^{-1}) were carried out in the gas phase and in solution, respectively. Each of these simulations was initialized with different random velocities drawn from a Maxwell-Boltzmann distribution at 300 K. In each gas phase run, 5 ns of simulation time were discarded as equilibration, followed by 10 ns of production, during which the restart files were saved every 1000th step. In the analogous solution simulations at constant volume, 0.5 ns were discarded as equilibration. During the subsequent 1 ns production phase, restart files were saved every 1000th step. Thus, during the cumulative simulation length of 8 ns (solution)/80 ns (gas phase), 8000 (solution)/80,000 (gas phase) restart files were saved. These served as the pool of configurations sampled in the canonical ensemble, from which non-equilibrium switching simulations to the high (SQM(/MM)) level of theory were started.

The solute molecules were always fully flexible; the TIP3 waters were held rigid using SHAKE [66]. In the gas phase, the nonbonded interactions were not truncated ("infinite" cutoff radius). In the solution simulations, Lennard-Jones interactions were smoothly switched off between 10 and 12 Å, and electrostatic interactions were computed with the particle-mesh-Ewald method [67] ($\kappa = 0.34$ $Å^{-1}$, spline interpolation to order 6, FFT grid size of 32).

The protocol just described was used both for the simulations at the MM level of theory, i.e., using the CGenFF force field, and for all simulations with a hybrid intermediate representation (force field with Mulliken-derived partial charges).

SQM(/MM) Equilibrium Simulations To calculate the reference values via Crook's equation, we also carried out simulations at the SQM(/MM) level of theory. Specifically, the solute was treated with the self-consistent-charge density-functional

tight-binding (SCC-DFTB) method, as implemented in CHARMM [68], using the 3ob-3-1 [69, 70, 71, 72] parameter set (`https://www.dftb.org/parameters/download/3ob/3ob-3-1-cc/`(accessed on 29 April 2023)). Water molecules were always treated classically. Analogous to the MM case just described, eight LD simulations started with different random initial velocities were carried out in the gas phase and in solution. The simulation length in all cases was 1 ns (1 million steps with a timestep of 1 fs). Restart files were written every 1000th step in solution and every 100th step in the gas phase, thus resulting in a total of 800 (solution)/8000 (gas phase) restart files generated during a cumulative simulation length of 8 ns. Nonbonded interactions were treated identically, as in the MM case.

Non-Equilibrium Work Simulations

From each set of restart files written during the equilibrium simulations at the MM, MULL, and SQM(/MM) levels of theory, every 40th was selected as the starting point for a NEW switch to the respective other level of theory. Thus, 200 switches per molecule were carried out to compute the free energy difference between two levels of theory. In all switching simulations, we used a timestep of 1 fs. Unless otherwise noted, the switching length was 2000 fs (2000 steps), both for linear and stepwise linear switching paths.

Calculation of $\Delta A_{Xsolv}^{\mathbf{MM} \leftrightarrow \mathbf{MULL}}$

As described in Section II.4.3, to close the thermodynamic cycle, one needs to compute $\Delta A_{Xsolv}^{\mathrm{MM} \leftrightarrow \mathrm{MULL}}$. Since the only difference between the MM and $MULL$ descriptions of a system are the partial charges of the solute, $\{q_i\}$, intermediate states can be set up easily according to $q_i(\lambda) = (1-\lambda)q_i^{MM} + \lambda q_i^{MULL}$. To ensure an overlap between neighboring states, we used three intermediate states ($\lambda = 0.25,\ 0.5,\ 0.75$) in addition to the MM ($\lambda = 0$) and MULL ($\lambda = 1$) end states. All simulation settings were the same as described in the force field-based equilibrium simulations; the production length per state was 1 ns. At each λ state, 1000 coordinate sets were saved and the energies were reevaluated at the neighboring states. The free energy difference between the neighboring states was computed with BAR; these were summed up to give $\Delta A_{Xsolv}^{\mathrm{MM} \leftrightarrow \mathrm{MULL}}$. All calculations were repeated eight times, starting from initial random velocities. The standard deviation obtained from these eight repetitions were used to estimate the standard error via Gaussian error propagation. All MM calculations were carried out with the OpenMM [73] GPU acceleration available in CHARMM (`https://academiccharmm.org/documentation/version/c47b1/openmm`(accessed on 29 April 2023)).

II.5 Conclusions

In summary, the reliable calculation of free energy differences between the MM and SQM/MM levels of theory in aqueous solution require either switching lengths of at least 5 ps, or the use of an appropriate intermediate charge state. In terms of aiding convergence, the use of intermediate states is not new. For example, the internal degrees of freedom of a molecule or region that should be described at two levels of theory can be made more similar to the target high-level representation via force matching [33, 29, 74]. Thus, the use of intermediate charge states plays the same role for electrostatic interactions between the core region (the region changed from e.g., MM to SQM) and the environment (the region always described at the low level of theory), as force-matched parameters play for the bonded energy terms of the core region. Since intermediate charge states can be rationally designed without too much computational effort, the poor convergence of MM→SQM/MM simulations due to slow solvent reorientation can be easily circumvented.

II.6 Abbreviations

MD	Molecular Dynamics
MM	Molecular Mechanics
(S)QM	(Semi-empirical-)Quantum Mechanics
(S)QM/MM	(Semi-empirical-)Quantum Mechanical Molecular Mechanical hybrid methods
FES	Free energy simulation
FEP	Free energy perturbation
TI	Thermodynamic integration
BAR	Bennett's acceptance ratio
NEW	Non-equilibrium work methods
JAR	Jarzynski's Equation
CRO	Crook's Equation

Supporting Information

The following can be found in supporting information (Appendix B): Figure B.1: $\delta\Delta A$ including error bars; Figure B.2: $\delta\Delta A$, all points; Figure B.3: $\delta\Delta A$, all points, including error bars; Figure B.4: RMSD[q]; Figure B.5: $\Delta\mu$[D]; Figure B.6: θ_{SQM}[°]; Figure B.7: 3-t2 charge analysis; Figure B.8: 4-t2 charge analysis; Figure B.9: 5-t2 charge analysis; Figure B.10: 6-t2 charge analysis; Figure B.11: 8-t2 charge analysis;

Figure B.12: 4-t2 TIP3P relaxation; Figure B.13: 6-t2 TIP3P relaxation; Figure B.14: 8-t2 TIP3P relaxation; Table B.1: Additional information on model compounds; Table B.2: Summary table of all $\delta\Delta A$ results in solution; Table B.3: Detailed results for N_{switch} = 2000 fs $N_{replicate}$ = 200 $\Delta A_{gas}^{MM\rightarrow SQM}$; Table B.4: Detailed results for N_{switch} = 2000 fs $N_{replicate}$ = 200 $\Delta A_{Xsolv}^{MM\rightarrow SQM}$ in solution; Table B.5: Detailed results for N_{switch} = 5000 fs $N_{replicate}$ = 200 $\Delta A_{Xsolv}^{MM\rightarrow SQM}$ in solution; Table B.6: Detailed results for N_{switch} = 2000 fs $N_{replicate}$ = 200 $\Delta A_{Xsolv}^{MULL(gas)\rightarrow SQM}$ in solution; Table B.7: Detailed results for energy correction term via FEP $\Delta A_{Xsolv}^{MULL(gas)\rightarrow SQM}$ in solution; Table B.8: Detailed results for N_{switch} = 2000 fs $N_{replicate}$ = 200 $\Delta A_{Xsolv}^{MULL(solv)\rightarrow SQM}$ in solution; Table B.9: Detailed results for energy correction term via FEP $\Delta A_{Xsolv}^{MULL(solv)\rightarrow SQM}$ in solution; Table B.10: Detailed results for N_{switch} = 2000 fs $N_{replicate}$ = 200 $\Delta A_{Xsolv}^{MULL(solv*)\rightarrow SQM}$ in solution; Table B.11: Detailed results for energy correction term via FEP $\Delta A_{Xsolv}^{MULL(solv*)\rightarrow SQM}$ in solution; Table B.12: Mulliken Charge Analysis; Table B.13: TIP3P Water Relaxation. Refs. [75, 76] are cited in the Supplementary Materials.

Acknowledgments

This research was funded by the National Institute of General Medical Sciences, grant number 1R01GM129519-01 (HLW), and by the Austrian Science Fund (FWF), grant number P-31024 (SB). The APC for this article were also covered by the Austrian Science Fund (FWF). We thank Christian Schröder for helping with the calculations of Stoke shift relaxations.

Bibliography

[1] P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, "Free energy calculations by computer simulation," *Science*, vol. 236, pp. 564–568, may 1987.

[2] J. Gao, K. Kuczera, B. Tidor, and M. Karplus, "Hidden thermodynamics of mutant proteins: A molecular dynamics analysis," *Science*, vol. 244, pp. 1069–1072, jun 1989.

[3] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, "Alchemical free energy methods for drug discovery: progress and challenges," *Curr. Opin. Struct. Biol.*, vol. 21, no. 2, pp. 150–160, 2011.

[4] E. P. Barros, B. Ries, L. Böselt, C. Champion, and S. Riniker, "Recent developments in multiscale free energy simulations," *Curr. Opin. Struct. Biol.*, vol. 72, pp. 55–62, 2022.

[5] A. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, *et al.*, "Best practices for alchemical free energy calculations [article v1.0]," *Living journal of computational molecular science*, vol. 2, no. 1, 2020.

[6] C. E. Schindler, H. Baumann, A. Blum, D. Boöse, H.-P. Buchstaller, L. Burgdorf, D. Cappel, E. Chekler, P. Czodrowski, D. Dorsch, *et al.*, "Large-scale assessment of binding free energy calculations in active drug discovery projects," *J. Chem. Inf. Model.*, vol. 60, no. 11, pp. 5457–5474, 2020.

[7] Z. Cournia, B. Allen, and W. Sherman, "Relative binding free energy calculations in drug discovery: Recent advances and practical considerations," *J. Chem. Inf. Model.*, vol. 57, pp. 2911–2937, dec 2017.

[8] V. Limongelli, "Ligand binding free energy and kinetics calculation in 2020," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 10, no. 4, p. e1455, 2020.

[9] S. Maier, B. Thapa, J. Erickson, and K. Raghavachari, "Comparative assessment of qm-based and mm-based models for prediction of protein–ligand binding affinity trends," *Phys. Chem. Chem. Phys.*, 2022.

[10] W. Yang, Y. Q. Gao, Q. Cui, J. Ma, and M. Karplus, "The missing link between thermodynamics and structure in f_1-ATPase," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 874–879, jan 2003.

[11] A. J. Mulholland, "Chemical accuracy in qm/mm calculations on enzyme-catalysed reactions," *Chem. Cent. J.*, vol. 1, no. 1, pp. 1–5, 2007.

[12] B. Lev, B. Roux, and S. Y. Noskov, "Relative free energies for hydration of monovalent ions from qm and qm/mm simulations," *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 4165–4175, 2013.

[13] S. Riahi, B. Roux, and C. N. Rowley, "Qm/mm molecular dynamics simulations of the hydration of mg (ii) and zn (ii) ions," *Can. J. Chem.*, vol. 91, no. 7, pp. 552–558, 2013.

[14] J. A. Delgado, V. Wineman-Fisher, S. Pandit, and S. Varma, "Inclusion of high-field target data in AMOEBA's calibration improves predictions of protein–ion interactions," *J. Chem. Inf. Model.*, vol. 62, pp. 4713–4726, sep 2022.

[15] W. Yang, Q. Cui, D. Min, and H. Li, "QM/MM Alchemical Free Energy Simulations: Challenges and Recent Developments," *Annu. Rep. Comput. Chem.*, vol. 6, pp. 51–62, 2010.

[16] J. Gao, "Absolute free energy of solvation from monte carlo simulations using combined quantum and molecular mechanical potentials," *J. Phys. Chem.*, vol. 96, pp. 537–540, jan 1992.

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

[18] Q. Cui, T. Pal, and L. Xie, "Biomolecular QM/MM simulations: What are some of the "burning issues"?," *J. Phys. Chem. B*, vol. 125, pp. 689–702, jan 2021.

[19] P. S. Hudson, F. Aviat, R. Meana-Pañeda, L. Warrensford, B. C. Pollard, S. Prasad, M. R. Jones, H. L. Woodcock, and B. R. Brooks, "Obtaining QM/MM binding free energies in the SAMPL8 drugs of abuse challenge: indirect approaches," *J. Comput.-Aided Mol. Des.*, pp. 1–15, 2022.

[20] X. Wang, Q. He, and Z. Sun, "BAR-based multi-dimensional nonequilibrium pulling for indirect construction of a QM/MM free energy landscape," *Phys. Chem. Chem. Phys.*, vol. 21, no. 12, pp. 6672–6688, 2019.

[21] Z. Sun and Z. Liu, "BAR-based multi-dimensional nonequilibrium pulling for indirect construction of QM/MM free energy landscapes: Varying the QM region," *Advanced Theory and Simulations*, vol. 4, no. 10, p. 2100185, 2021.

[22] J.-N. Wang, W. Liu, P. Li, Y. Mo, W. Hu, J. Zheng, X. Pan, Y. Shao, and Y. Mei, "Accelerated computation of free energy profile at ab initio quantum mechanical/molecular mechanics accuracy via a semiempirical reference potential. 4. adaptive QM/MM," *J. Chem. Theory Comput.*, vol. 17, no. 3, pp. 1318–1325, 2021. PMID: 33593057.

[23] P. S. Hudson, K. Han, H. L. Woodcock, and B. R. Brooks, "Force matching as a stepping stone to qm/mm cb [8] host/guest binding free energies: a sampl6 cautionary tale," *J. Comput. -Aided Mol. Des.*, vol. 32, no. 10, pp. 983–999, 2018.

[24] G. König, F. C. Pickard, Y. Mei, and B. R. Brooks, "Predicting hydration free energies with a hybrid qm/mm approach: an evaluation of implicit and explicit solvation models in sampl4," *J. Comput. -Aided Mol. Des.*, vol. 28, pp. 245–257, 2014.

[25] G. König, B. R. Brooks, W. Thiel, and D. M. York, "On the convergence of multi-scale free energy simulations," *Mol. Simul.*, vol. 44, no. 13-14, pp. 1062–1081, 2018.

[26] G. König and B. R. Brooks, "Predicting binding affinities of host-guest systems in the sampl3 blind challenge: the performance of relative free energy calculations," *J. Comput. -Aided Mol. Des.*, vol. 26, pp. 543–550, 2012.

[27] J. Zeng, Y. Tao, T. J. Giese, and D. M. York, "Modern semiempirical electronic structure methods and machine learning potentials for drug discovery: Conformers, tautomers, and protonation states," *J. Chem. Phys.*, vol. 158, p. 124110, mar 2023.

[28] J. Heimdal and U. Ryde, "Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations," *Phys. Chem. Chem. Phys.*, vol. 14, no. 36, pp. 12592–12604, 2012.

[29] T. J. Giese and D. M. York, "Development of a robust indirect approach for MM→QM free energy calculations that combines force-matched reference potential and bennett's acceptance ratio methods," *J. Chem. Theory Comput.*, vol. 15, no. 10, pp. 5543–5562, 2019.

[30] P. S. Hudson, H. L. Woodcock, and S. Boresch, "Use of Nonequilibrium Work Methods to Compute Free Energy Differences Between Molecular Mechanical and Quantum Mechanical Representations of Molecular Systems," *J. Phys. Chem. Lett.*, vol. 6, no. 23, pp. 4850–4856, 2015.

[31] F. L. Kearns, P. S. Hudson, H. L. Woodcock, and S. Boresch, "Computing Converged Free Energy Differences Between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1376–1388, 2017.

[32] S. Boresch and H. L. Woodcock, "Convergence of single-step free energy perturbation," *Mol. Phys.*, vol. 115, no. 9-12, pp. 1200–1213, 2017.

[33] P. S. Hudson, S. Boresch, D. M. Rogers, and H. L. Woodcock, "Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching," *J. Chem. Theory Comput.*, vol. 14, no. 12, pp. 6327–6335, 2018.

[34] F. L. Kearns, L. Warrensford, S. Boresch, and H. L. Woodcock, "The Good, the Bad, and the Ugly: "HiPen", a New Dataset for Validating (S)QM/MM Free Energy Simulations," *Molecules*, vol. 24, no. 4, 2019.

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

[36] A. Schöller, F. Kearns, H. L. Woodcock, and S. Boresch, "Optimizing the calculation of free energy differences in nonequilibrium work SQM/MM switching simulations," *J. Phys. Chem. B*, vol. 126, no. 15, pp. 2798–2811, 2022.

[37] S. Ito and Q. Cui, "Multi-level free energy simulation with a staged transformation approach," *J. Chem. Phys.*, vol. 153, p. 044115, jul 2020.

[38] M. Maroncelli and G. R. Fleming, "Computer simulation of the dynamics of aqueous solvation," *J. Chem. Phys.*, vol. 89, no. 8, pp. 5044–5069, 1988.

[39] M. Maroncelli, "The dynamics of solvation in polar liquids," *J. Mol. Liq.*, vol. 57, pp. 1–37, 1993.

[40] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, no. 14, p. 2690, 1997.

[41] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, no. 3, p. 2361, 2000.

[42] C. H. Bennett, "Efficient estimation of free energy differences from monte carlo data," *J. Comput. Phys.*, vol. 22, no. 2, pp. 245–268, 1976.

[43] A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge model 5: An extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases.," *J. Chem. Theory Comput.*, vol. 8, pp. 527–541, feb 2012.

[44] C. Dellago and G. Hummer, "Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics," *Entropy*, vol. 16, no. 1, pp. 41–61, 2014.

[45] A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free-energy calculations," *J. Phys. Chem. B*, vol. 114, pp. 10235–10253, jul 2010.

[46] H. L. Woodcock, B. T. Miller, M. Hodoscek, A. Okur, J. D. Larkin, J. W. Ponder, and B. R. Brooks, "Mscale: a general utility for multiscale modeling," *J. Chem. Theory Comput.*, vol. 7, no. 4, pp. 1208–1219, 2011.

[47] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, "CHARMM: the biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.

[48] J. E. Hunter, W. P. Reinhardt, and T. F. Davis, "A finite-time variational method for determining optimal paths and obtaining bounds on free energy changes from computer simulations," *J. Chem. Phys.*, vol. 99, pp. 6856–6864, nov 1993.

[49] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, and P. J. Taylor, "The SAMPL2 Blind Prediction Challenge: Introduction and Overview," *J. Comput.-Aided Mol. Des.*, vol. 24, no. 4, pp. 259–279, 2010.

[50] O. Wahl and T. Sander, "Tautobase: an open tautomer database," *J. Chem. Inf. Model.*, vol. 60, no. 3, pp. 1085–1089, 2020.

[51] M. Wieder, J. Fass, and J. D. Chodera, "Fitting quantum machine learning potentials to experimental free energy data: predicting tautomer ratios in solution," *Chem. Sci.*, vol. 12, no. 34, pp. 11364–11381, 2021.

[52] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Phys. Rev. B*, vol. 58, no. 11, p. 7260, 1998.

[53] M. Elstner, "The SCC-DFTB method and its application to biological systems," *Theor. Chem. Acc.*, vol. 116, no. 1, pp. 316–325, 2006.

[54] D. A. Pearlman and P. A. Kollman, "The lag between the hamiltonian and the system configuration in free energy perturbation calculations," *J. Chem. Phys.*, vol. 91, no. 12, pp. 7831–7839, 1989.

[55] J. Hermans, "Simple analysis of noise and hysteresis in (slow-growth) free energy simulations," *J. Phys. Chem.*, vol. 95, no. 23, pp. 9029–9032, 1991.

[56] C. Jarzynski, "Rare events and the convergence of exponentially averaged work values," *Phys. Rev. E*, vol. 73, no. 4, p. 046105, 2006.

[57] E. Heid, S. Harringer, and C. Schröder, "The small impact of various partial charge distributions in ground and excited state on the computational stokes shift of 1-methyl-6-oxyquinolinium betaine in diverse water models," *J. Chem. Phys.*, vol. 145, p. 164506, oct 2016.

[58] C. Schröder and E. Heid, "Chapter four - computational solvation dynamics: Implementation, application, and validation," vol. 16 of *Annual Reports in Computational Chemistry*, pp. 93–154, Elsevier, 2020.

[59] S. Jo, T. Kim, V. G. Iyer, and W. Im, "Charmm-gui: a web-based graphical user interface for charmm," *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, 2008.

[60] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, *et al.*, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, 2016.

[61] K. Vanommeslaeghe and A. D. MacKerell Jr, "Automation of the charmm general force field (cgenff) i: bond perception and atom typing," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3144–3154, 2012.

[62] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell Jr, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3155–3168, 2012.

[63] W. Yu, X. He, K. Vanommeslaeghe, and A. D. MacKerell Jr, "Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations," *J. Comput. Chem.*, vol. 33, no. 31, pp. 2451–2468, 2012.

[64] E. Neria, S. Fischer, and M. Karplus, "Simulation of activation free energies in molecular systems," *J. Chem. Phys.*, vol. 105, p. 1902, 1996.

[65] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, "Constant pressure molecular dynamics simulation: The langevin piston method," *J. Chem. Phys.*, vol. 103, no. 11, pp. 4613–4621, 1995.

[66] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, pp. 327–341, mar 1977.

[67] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, 1995.

[68] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, "A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method," *J. Phys. Chem. B*, vol. 105, no. 2, pp. 569–585, 2001.

[69] M. Gaus, A. Goez, and M. Elstner, "Parametrization and Benchmark of DFTB3 for Organic Molecules," *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 338–354, 2013. PMID: 26589037.

[70] M. Gaus, X. Lu, M. Elstner, and Q. Cui, "Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications," *J. Chem. Theory Comput.*, vol. 10, no. 4, pp. 1518–1537, 2014. PMID: 24803865.

[71] X. Lu, M. Gaus, M. Elstner, and Q. Cui, "Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications," *J. Phys. Chem. B*, vol. 119, no. 3, pp. 1062–1082, 2015. PMID: 25178644.

[72] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, "Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems," *J. Chem. Theory Comput.*, vol. 11, no. 1, pp. 332–342, 2015.

[73] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLoS Comput. Biol.*, vol. 13, no. 7, p. e1005659, 2017.

[74] X. Pan, P. Li, J. Ho, J. Pu, Y. Mei, and Y. Shao, "Accelerated computation of free energy profile at ab initio quantum mechanical/molecular mechanical

accuracy via a semi-empirical reference potential. ii. recalibrating semi-empirical parameters with force matching," *Phys. Chem. Chem. Phys.*, vol. 21, no. 37, pp. 20595–20605, 2019.

[75] R. H. Wood, W. C. F. Muhlbauer, and P. T. Thompson, "Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis," *J. Phys. Chem.*, vol. 95, no. 17, pp. 6670–6675, 1991.

[76] D. Wu and D. A. Kofke, "Model for small-sample bias of free-energy calculations applied to gaussian-distributed nonequilibrium work measurements," *J. Chem. Phys.*, vol. 121, no. 18, pp. 8742–8747, 2004.

Appendices

List of Tables

List of Figures

A. Supplementary Material of Publication I

Supplementary information to Optimizing the Calculation of Free Energy Differences in Nonequilibrium Work SQM/MM Switching Simulations

Tab. A.1. ZINC ID [71], ID used in Ref. [35] and in this manuscript, total number of atoms (N_{total}), the total number of heavy atoms (N_{heavy}) and free energy offset for each compound investigated. This offset has to be subtracted from the respective $\Delta A^{MM \to SQM}$ ($\Delta A^{SQM \to MM}$) listed in Tables A.3–A.6 below.

ZINC-ID	ID	N_{tot}/N_{heavy}	Offset (kcal/mol)
00077329	2	16/10	15000
00079729	3	18/13	17000
00086442	4	21/12	19000
00107550	7	21/11	16000
00133435	10	34/22	28000
00138607	11	36/20	29000
00140610	12	20/12	17000
00164361	13	23/14	20000
00167648	14	44/26	35000
00169358	15	26/16	22000
01867000	17	32/18	22000
06568023	21	30/18	25000

Tab. A.2. MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) of results obtained from forward switches compared to Reference result (CRO). These are the values resulting from the short equilibration protocols used initially; this table corresponds to Fig.I.3 of the main manuscript.

	MAD [kcal/mol]	RMSE [kcal/mol]
NSWI200	0.80	1.79
NSWI500	0.56	1.36
NSWI1000	0.49	1.07
NSWI2000	0.20	0.36

Convergence criteria reported in Tables A.3–A.6: In line with earlier work [35], we computed several convergence criteria for each free energy difference $\Delta A^{MM \rightarrow SQM}$. As described in the main manuscript, raw data for each ΔA were generated through eight independent series of calculations. We consider each set of raw data as a block and calculated ΔA_i for each of these blocks. Thus, we can compute a "hysteresis" Hyst(eresis) $= \Delta A - 1/8 \sum_{i=1}^{8} \Delta A_i$; cf. Ref. [72]. Here, ΔA is the free energy computed using Jarzynski's equation from *all* available data, combining the work values of the eight series of calculations. Each ΔA_i is the free energy computed from the work values of one of the eight series of calculations.

Further, we report the one-sided Π values for the forward and backward free energy differences; see Ref. [73] for the definition of this quantity and the rationale behind it. Π values > 0.5 indicate that the W distribution is based on sufficient and unbiased sampling. We report two variants for Π, one evaluated with the initial guess for ΔA calculated by CRO (Π_{CRO}), the other by JAR (Π_{JAR}). Here (cf. main manuscript) CRO refers to Crook's equation [40], and JAR to Jarzynski's equation [32].

Contents of Tables A.3–A.6: $\Delta A_{gas}^{MM \rightarrow SQM}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma \Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations. For compounds **2**, **7**, **10**, **12** and **21** entries in italics are for the original protocol, entries in bold indicate the results obtained with longer equilibration after the random dihedral assignment.

Tab. A.3. Detailed results for N_{switch}=200 fs $N_{replicate}$=8000

	JAR(fw)							JAR(bw)							CRO		Overlap (%)
Paper-ID	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W
2	*-258.08*	*0.28*	*0.59*	*-256.35*	*1.33*	*1.36*	*1.27*	*257.39*	*0.11*	*0.39*	*258.82*	*2.11*	*1.99*	*1.49*	*-257.44*	*0.02*	*26.37*
2	**-257.84**	**0.28**	**0.55**	**-256.02**	**1.26**	**0.55**	**1.20**	**257.17**	**0.92**	**1.78**	**259.91**	**2.76**	**2.13**	**0.64**	**-257.37**	**0.02**	**20.35**
3	-414.30	0.01	0.10	-412.75	1.50	2.22	1.39	413.90	0.02	0.15	415.05	1.00	1.32	1.72	-414.08	0.02	22.62
4	-256.52	0.00	0.03	-256.00	0.79	2.59	2.36	256.04	0.00	0.03	256.58	0.86	2.81	2.34	-256.28	0.01	58.07
7	*-994.14*	*0.39*	*0.75*	*-992.56*	*1.08*	*0.49*	*1.37*	*993.21*	*1.27*	*2.00*	*996.10*	*3.00*	*2.88*	*0.57*	*-993.57*	*0.02*	*26.04*
7	**-993.20**	**0.19**	**0.43**	**-992.05**	**0.96**	**1.59**	**1.71**	**992.59**	**0.03**	**0.21**	**993.65**	**1.51**	**2.61**	**1.79**	**-992.71**	**0.01**	**35.29**
10	*-338.40*	*0.28*	*0.59*	*-335.73*	*1.72*	*0.92*	*0.69*	*337.79*	*0.06*	*0.26*	*340.23*	*1.71*	*0.95*	*0.82*	*-337.97*	*0.03*	*11.04*
10	**-338.30**	**0.20**	**0.53**	**-335.67**	**1.67**	**0.88**	**0.71**	**337.91**	**0.06**	**0.26**	**340.22**	**1.70**	**0.96**	**0.89**	**-337.90**	**0.03**	**10.81**
11	-462.22	0.07	0.33	-459.19	3.09	1.73	0.49	461.95	0.00	0.12	463.65	1.64	0.47	1.29	-462.00	0.02	21.64
12	*-76.70*	*0.06*	*0.19*	*-75.51*	*1.06*	*1.73*	*1.68*	*75.78*	*0.03*	*0.12*	*77.12*	*1.38*	*2.40*	*1.56*	*-76.22*	*0.00*	*33.82*
12	**-76.68**	**0.04**	**0.22**	**-75.50**	**1.06**	**1.76**	**1.67**	**75.82**	**0.05**	**0.08**	**77.09**	**1.37**	**2.38**	**1.61**	**-76.21**	**0.02**	**34.37**
13	-567.91	0.00	0.04	-567.44	0.75	1.73	1.68	567.32	0.00	0.04	567.81	0.78	2.40	1.56	-567.62	0.01	68.17
14	-82.73	0.04	0.23	-81.35	1.20	1.41	1.53	82.36	0.01	0.10	83.96	1.48	1.51	1.36	-82.57	0.02	20.31
15	-407.80	0.00	0.06	-407.38	0.64	2.24	2.50	407.49	0.00	0.02	407.99	0.88	3.19	2.38	-407.64	0.01	52.92
17	-673.51	0.00	0.07	-672.91	0.81	2.31	2.26	673.21	0.00	0.05	673.85	0.90	2.54	2.22	-673.36	0.01	41.08
21	*-75.12*	*4.22*	*3.09*	*-68.27*	*1.61*	*0.92*	*-1.12*	*68.80*	*0.01*	*0.07*	*70.75*	*2.14*	*2.88*	*1.12*	*-69.08*	*0.02*	*24.14*
21	**-69.00**	**0.01**	**0.11**	**-68.00**	**0.95**	**0.94**	**1.84**	**68.78**	**0.01**	**0.08**	**70.74**	**2.16**	**2.48**	**1.12**	**-68.99**	**0.03**	**23.31**

Tab. A.4. Detailed results for N_{switch}=500 fs $N_{replicate}$=3200

	JAR(fw)							JAR(bw)							CRO		Overlap (%)
Paper-ID	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W
2	*-258.14*	*0.37*	*0.55*	*-256.97*	*0.83*	*1.33*	*1.46*	*257.43*	*0.02*	*0.20*	*258.36*	*1.71*	*2.75*	*1.68*	*-257.48*	*0.02*	*37.72*
2	**-257.38**	**0.04**	**0.21**	**-256.73**	**0.77**	**0.69**	**1.96**	**257.28**	**0.79**	**1.61**	**259.02**	**2.25**	**3.13**	**1.02**	**-257.37**	**0.02**	**33.93**
3	-414.23	0.01	0.10	-413.23	1.27	2.44	1.61	413.95	0.06	0.26	414.72	0.77	1.42	1.84	-414.11	0.03	31.41
4	-256.39	0.00	0.07	-256.12	0.54	2.59	2.49	256.19	0.00	0.03	256.45	0.59	2.83	2.51	-256.28	0.02	64.48
7	*-993.87*	*0.23*	*0.60*	*-992.91*	*0.83*	*0.46*	*1.64*	*993.28*	*1.28*	*2.02*	*995.68*	*2.79*	*3.35*	*0.60*	*-993.59*	*0.02*	*32.58*
7	**-992.79**	**0.01**	**0.12**	**-992.35**	**0.67**	**1.60**	**2.23**	**992.64**	**0.03**	**0.21**	**993.31**	**1.21**	**2.96**	**1.94**	**-992.72**	**0.02**	**46.86**
10	*-338.38*	*0.25*	*0.58*	*-336.83*	*1.20*	*1.21*	*1.16*	*337.88*	*0.01*	*0.10*	*339.27*	*1.47*	*1.66*	*1.28*	*-337.94*	*0.03*	*22.51*
10	**-337.92**	**0.03**	**0.20**	**-336.75**	**1.13**	**1.17**	**1.46**	**337.73**	**0.01**	**0.10**	**339.30**	**1.51**	**1.66**	**1.14**	**-337.88**	**0.03**	**21.59**
11	-462.11	0.06	0.19	-459.72	2.90	2.27	0.61	461.99	0.00	0.17	463.06	1.17	0.46	1.55	-461.99	0.03	28.04
12	*-76.37*	*0.00*	*0.05*	*-75.87*	*0.75*	*2.13*	*2.14*	*76.03*	*0.01*	*0.03*	*76.60*	*0.88*	*2.57*	*2.06*	*-76.20*	*0.02*	*50.00*
12	**-76.38**	**0.01**	**0.05**	**-75.87**	**0.74**	**2.11**	**2.13**	**76.04**	**0.01**	**0.06**	**76.60**	**0.88**	**2.58**	**2.07**	**-76.20**	**0.02**	**49.03**
13	-567.72	0.00	0.01	-567.52	0.49	2.13	2.14	567.48	0.00	0.02	567.70	0.51	2.57	2.06	-567.61	0.02	75.75
14	-82.65	0.00	0.06	-81.62	1.03	1.53	1.58	82.52	0.01	0.19	83.60	1.16	1.71	1.54	-82.58	0.03	22.59
15	-407.69	0.01	0.04	-407.48	0.48	2.42	2.58	407.57	0.00	0.03	407.82	0.59	2.98	2.54	-407.63	0.02	59.33
17	-673.41	0.00	0.03	-673.17	0.53	2.55	2.54	673.30	0.00	0.03	673.55	0.57	2.74	2.52	-673.36	0.02	58.35
21	*-73.64*	*0.88*	*1.01*	*-68.62*	*1.46*	*1.28*	*-0.67*	*68.91*	*0.00*	*0.03*	*70.06*	*1.54*	*3.66*	*1.48*	*-69.07*	*0.02*	*32.66*
21	**-69.05**	**0.01**	**0.12**	**-68.34**	**0.78**	**1.22**	**1.90**	**68.88**	**0.01**	**0.09**	**70.02**	**1.53**	**2.52**	**1.15**	**-68.98**	**0.02**	**30.68**

Tab. A.5. Detailed results for N_{switch}=1000 fs $N_{replicate}$=1600

	JAR(fw)							JAR(bw)							CRO		Overlap (%)
Paper-ID	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W
2	*-257.54*	*0.04*	*0.22*	*-257.12*	*0.61*	*1.20*	*2.07*	*257.46*	*0.12*	*0.39*	*258.20*	*1.57*	*3.06*	*1.68*	*-257.48*	*0.03*	*43.15*
2	**-257.49**	**0.08**	**0.32**	**-256.94**	**0.57**	**0.66**	**1.90**	**257.34**	**0.79**	**1.59**	**258.73**	**2.01**	**3.48**	**1.09**	**-257.39**	**0.04**	**39.58**
3	-414.18	0.00	0.05	-413.65	0.93	2.49	1.92	413.94	0.08	0.28	414.49	0.64	1.71	1.89	-414.13	0.03	42.57
4	-256.35	0.00	0.02	-256.21	0.41	2.55	2.55	256.25	0.00	0.02	256.40	0.45	2.82	2.54	-256.30	0.02	70.56
7	*-995.26*	*1.30*	*1.22*	*-993.03*	*0.74*	*0.41*	*0.52*	*993.28*	*1.26*	*1.95*	*995.47*	*2.69*	*3.44*	*0.54*	*-993.59*	*0.03*	*33.11*
7	**-992.84**	**0.03**	**0.21**	**-992.44**	**0.52**	**1.53**	**2.10**	**992.70**	**0.02**	**0.14**	**993.16**	**1.02**	**3.03**	**2.00**	**-992.71**	**0.03**	**51.53**
10	*-338.34*	*0.20*	*0.47*	*-337.35*	*0.89*	*1.35*	*1.42*	*337.87*	*0.01*	*0.08*	*338.74*	*1.24*	*2.19*	*1.54*	*-337.93*	*0.03*	*34.57*
10	**-337.91**	**0.03**	**0.18**	**-337.27**	**0.84**	**1.35**	**1.78**	**337.85**	**0.01**	**0.08**	**338.75**	**1.25**	**2.09**	**1.52**	**-337.90**	**0.03**	**33.74**
11	-462.14	0.06	0.28	-460.17	2.63	2.61	0.68	462.08	0.01	0.12	462.76	0.89	0.49	1.75	-462.01	0.04	34.43
12	*-76.31*	*0.00*	*0.03*	*-76.04*	*0.55*	*2.31*	*2.30*	*76.12*	*0.00*	*0.06*	*76.41*	*0.62*	*2.62*	*2.26*	*-76.21*	*0.03*	*62.42*
12	**-76.29**	**0.01**	**0.06**	**-76.02**	**0.54**	**2.28**	**2.31**	**76.14**	**0.01**	**0.07**	**76.44**	**0.62**	**2.56**	**2.25**	**-76.22**	**0.03**	**57.54**
13	-567.69	0.00	0.02	-567.58	0.37	2.31	2.30	567.55	0.00	0.04	567.66	0.37	2.62	2.26	-567.62	0.02	84.06
14	-82.50	0.04	0.23	-81.78	0.90	1.59	1.69	82.65	0.01	0.11	83.34	0.93	1.66	1.72	-82.55	0.03	25.32
15	-407.66	0.00	0.03	-407.52	0.38	2.49	2.58	407.59	0.00	0.04	407.73	0.43	2.84	2.56	-407.62	0.02	66.86
17	-673.39	0.00	0.03	-673.27	0.38	2.59	2.61	673.34	0.00	0.03	673.48	0.41	2.73	2.58	-673.37	0.02	67.30
21	*-72.31*	*2.14*	*1.81*	*-68.76*	*1.14*	*1.65*	*-0.20*	*68.93*	*0.00*	*0.06*	*69.63*	*1.14*	*3.74*	*1.72*	*-69.04*	*0.03*	*40.53*
21	**-69.19**	**0.09**	**0.31**	**-68.57**	**0.65**	**1.46**	**1.80**	**68.98**	**0.01**	**0.07**	**69.66**	**1.09**	**2.44**	**1.74**	**-69.01**	**0.03**	**37.09**

Tab. A.6. Detailed results for N_{switch}=2000 fs $N_{replicate}$=800

	JAR(fw)							JAR(bw)							CRO		Overlap (%)
Paper-ID	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W
2	*-258.05*	*0.36*	*0.69*	*-257.21*	*0.55*	*1.05*	*1.37*	*257.48*	*0.11*	*0.37*	*258.11*	*1.47*	*3.18*	*1.59*	*-257.48*	*0.04*	*46.51*
2	**-257.53**	**0.16**	**0.43**	**-257.03**	**0.42**	**0.57**	**1.76**	**257.36**	**0.65**	**1.40**	**258.58**	**1.90**	**3.61**	**1.03**	**-257.34**	**0.04**	**39.58**
3	-414.12	0.00	0.06	-413.87	0.59	2.38	2.14	414.07	0.02	0.17	414.33	0.48	1.98	2.11	-414.12	0.04	48.74
4	-256.32	0.00	0.04	-256.25	0.30	2.42	2.54	256.26	0.00	0.02	256.34	0.34	2.91	2.54	-256.29	0.03	79.22
7	*-993.91*	*0.34*	*0.71*	*-993.13*	*0.66*	*0.33*	*1.44*	*993.28*	*1.24*	*1.91*	*995.37*	*2.65*	*3.42*	*0.41*	*-993.61*	*0.04*	*33.62*
7	**-992.78**	**0.04**	**0.22**	**-992.51**	**0.39**	**1.33**	**2.11**	**992.71**	**0.02**	**0.17**	**993.11**	**0.98**	**3.19**	**1.88**	**-992.71**	**0.03**	**56.09**
10	*-338.34*	*0.19*	*0.48*	*-337.60*	*0.67*	*1.34*	*1.48*	*337.91*	*0.01*	*0.12*	*338.49*	*1.05*	*2.50*	*1.66*	*-337.94*	*0.04*	*43.62*
10	**-337.95**	**0.05**	**0.25**	**-337.51**	**0.61**	**1.36**	**1.83**	**337.89**	**0.01**	**0.12**	**338.46**	**1.06**	**2.37**	**1.67**	**-337.89**	**0.04**	**44.02**
11	-462.02	0.06	0.31	-460.62	2.29	2.82	0.88	461.18	0.55	0.72	462.54	0.72	0.54	0.92	-462.01	0.05	37.35
12	*-76.26*	*0.00*	*0.03*	*-76.13*	*0.38*	*2.42*	*2.40*	*76.18*	*0.00*	*0.03*	*76.32*	*0.42*	*2.56*	*2.36*	*-76.22*	*0.03*	*69.52*
12	**-76.24**	**0.01**	**0.04**	**-76.11**	**0.39**	**2.34**	**2.39**	**76.19**	**0.01**	**0.05**	**76.35**	**0.44**	**2.50**	**2.33**	**-76.22**	**0.03**	**65.29**
13	-567.66	0.00	0.02	-567.60	0.27	2.42	2.40	567.60	0.00	0.02	567.65	0.27	2.56	2.36	-567.62	0.03	86.77
14	-82.52	0.02	0.15	-81.94	0.80	1.65	1.66	82.62	0.01	0.12	83.14	0.77	1.63	1.73	-82.54	0.04	29.65
15	-407.65	0.00	0.05	-407.58	0.29	2.51	2.56	407.61	0.00	0.03	407.69	0.31	2.73	2.54	-407.63	0.03	75.41
17	-673.38	0.00	0.03	-673.31	0.28	2.58	2.58	673.35	0.00	0.04	673.42	0.30	2.66	2.56	-673.37	0.03	73.05
21	*-70.02*	*0.61*	*0.76*	*-68.79*	*0.74*	*2.08*	*1.02*	*68.99*	*0.01*	*0.10*	*69.37*	*0.79*	*3.61*	*1.93*	*-69.01*	*0.04*	*46.79*
21	**-69.00**	**0.01**	**0.09**	**-68.73**	**0.53**	**1.69**	**2.09**	**68.96**	**0.01**	**0.09**	**69.38**	**0.82**	**2.43**	**1.86**	**-69.01**	**0.04**	**47.98**

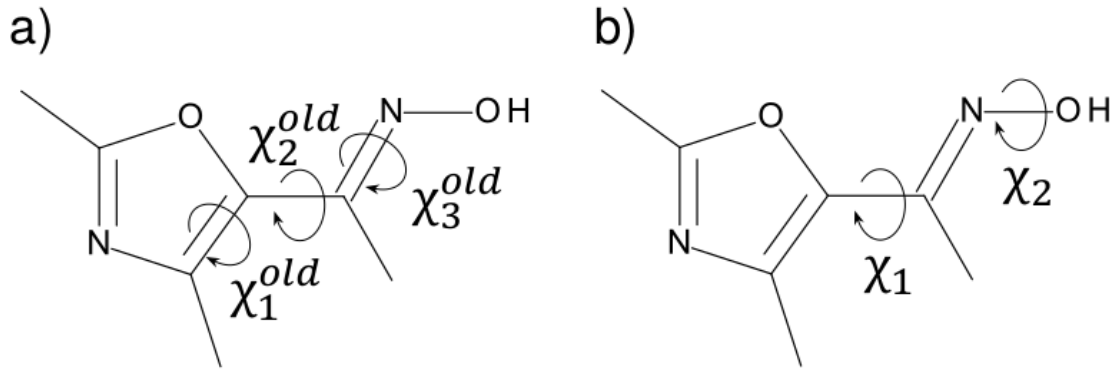


Fig. A.1. Changes in dihedral angle nomenclature for compound **7**; a) dihedral labels used in Ref. [35]; b) dihedral labels used in this work.

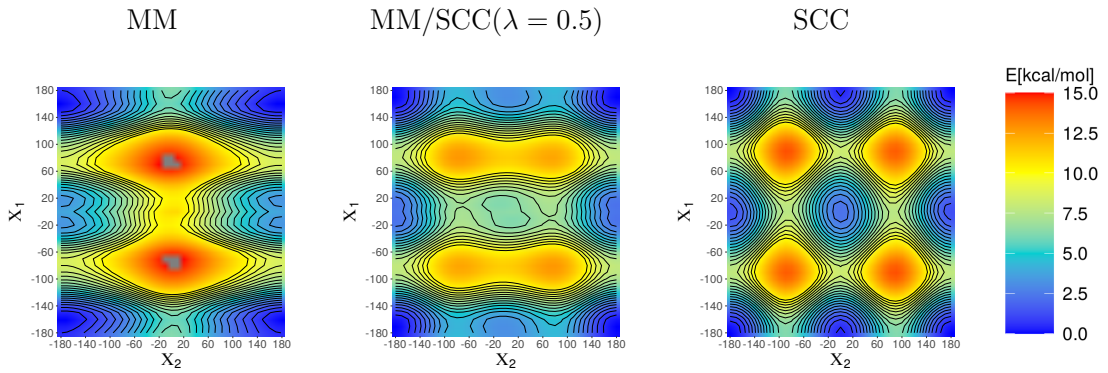


Fig. A.2. Potential energy scan as a function of χ_1 and χ_2 (cf. Fig. I.1 of the main manuscript and Fig. A.1 above) for compound **7**. Left: MM; middle MM/3ob at $\lambda = 0.5$; right: SQM (=3ob). All energies are in kcal/mol, the contour lines are plotted in 0.5 kcal/mol increments from 0 to 10 kcal/mol. Energies were rescaled so that the minimum energy is at 0 kcal/mol in all three plots.

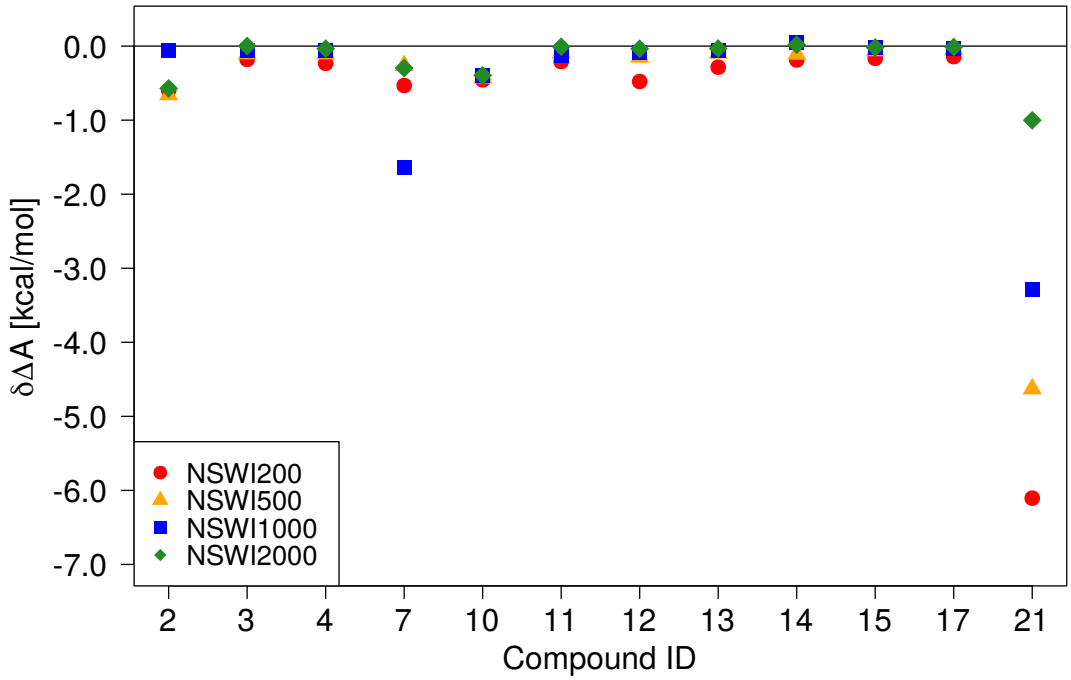


Fig. A.3. $\delta\Delta A$ for the results of the forward switching simulations (MM $\rightarrow$ SQM) using the original equilibration protocol as a function of switching length. Here all values, including those that were off-scale in Fig. I.3 of the manuscript, are shown.

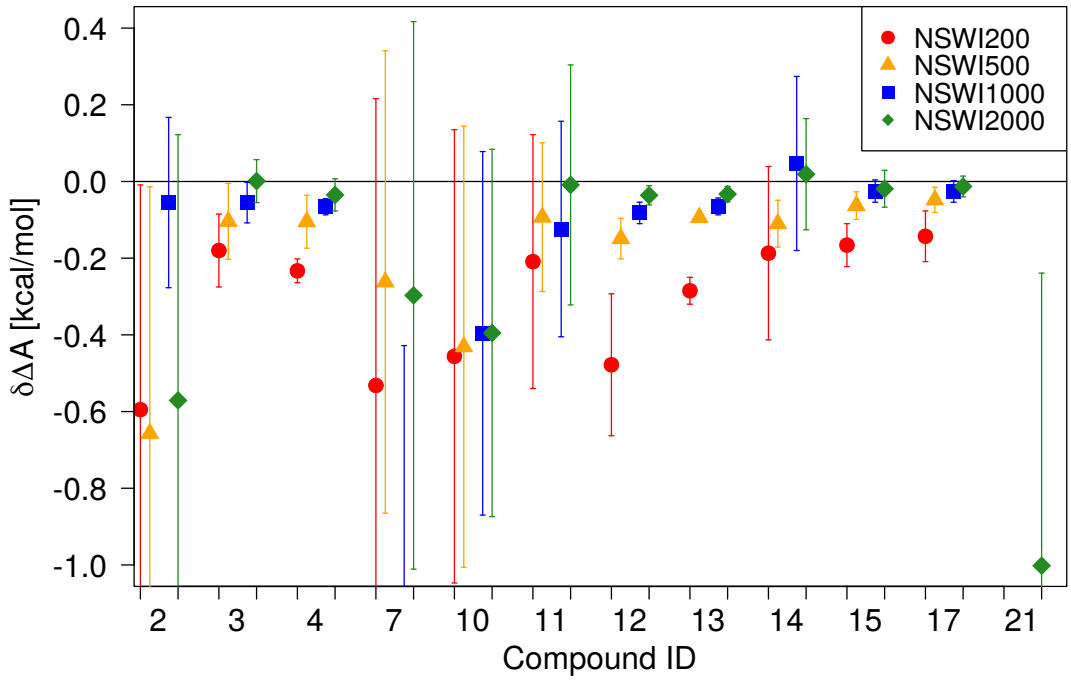


Fig. A.4. $\delta\Delta A$ as shown in Fig. I.3 of the main manuscript (original equilibration protocol) with error bars, based on $\sigma_{\Delta A_i}$ from Tables S4-S7.

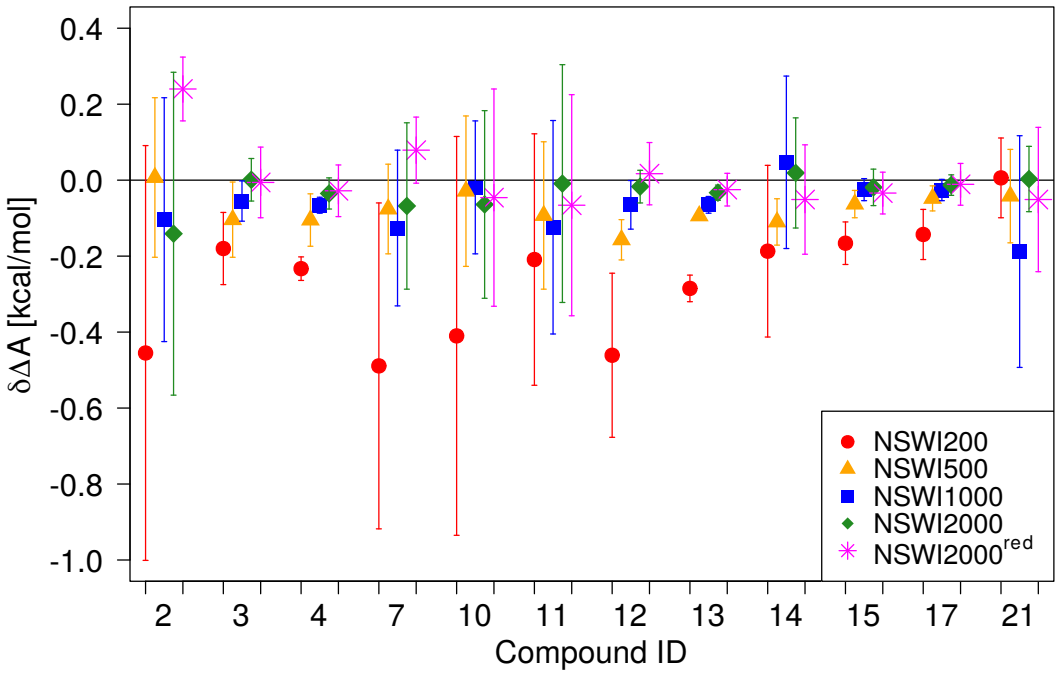


Fig. A.5. $\delta\Delta A$ as shown in Fig. I.7 of the main manuscript (longer equilibration for **2**, **7**, **10**, **12** and **21**) with error bars, based on $\sigma_{\Delta A_i}$ from Tables S4-S7.

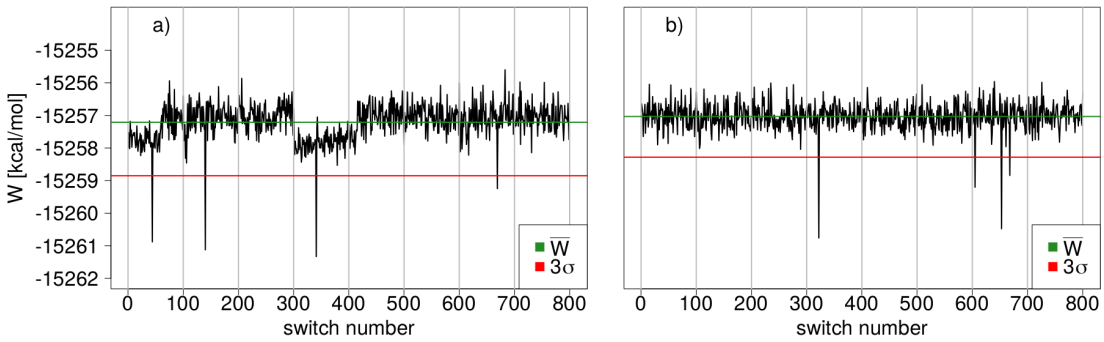


Fig. A.6. Quasi time series of compound **2** for a) the original protocol and for b) the modified equilibration protocol.

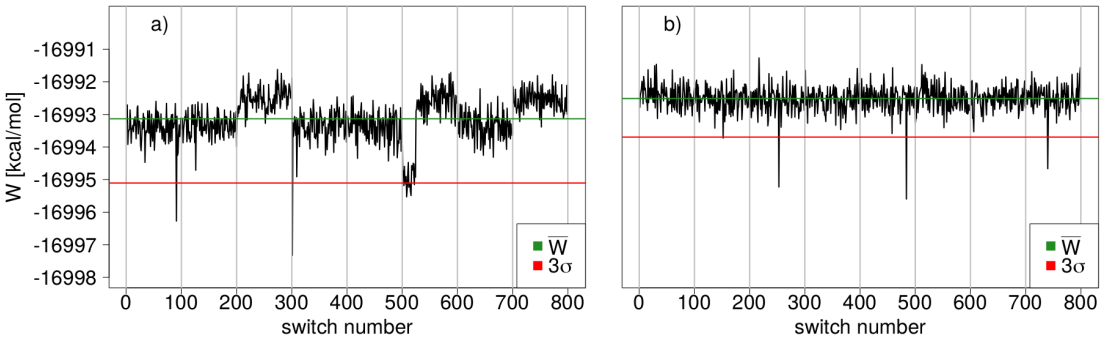


Fig. A.7. Quasi time series of compound **7** for a) the original protocol and for b) the modified equilibration protocol.

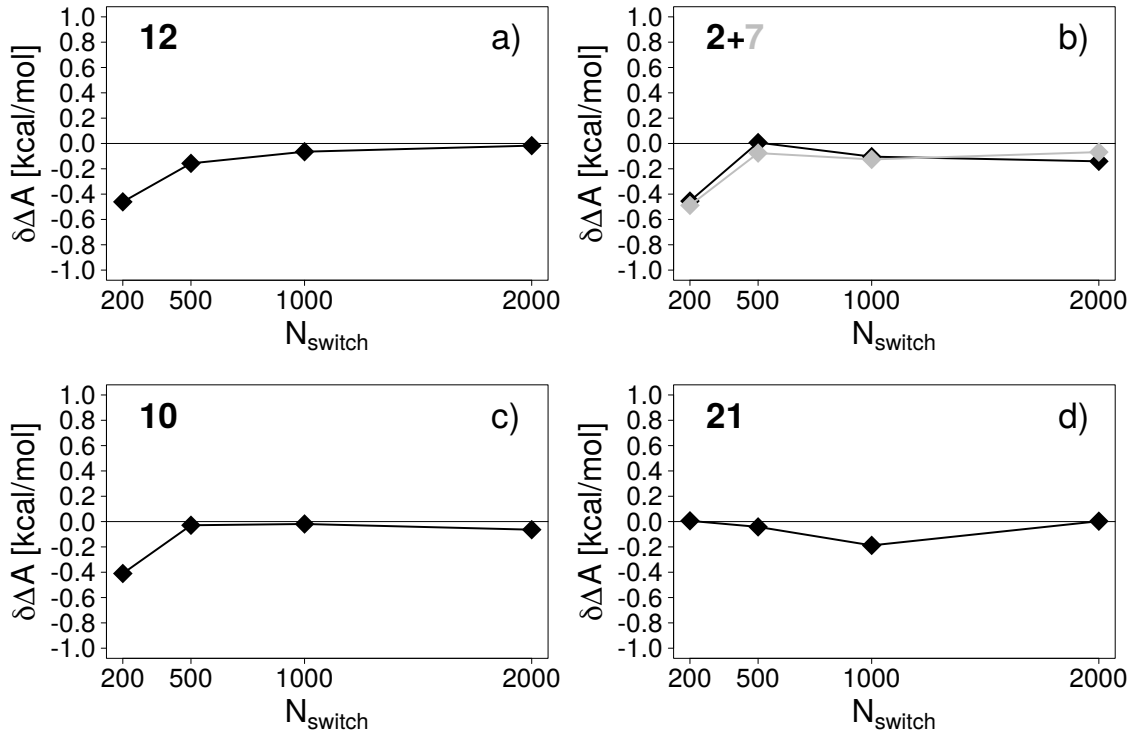


Fig. A.8. Detailed dependence of $\delta\Delta A$ on switching length for a) compounds **12** b) **2** and **7**, c)**10** and d) **21** with extended pre-production equilibration time. This is the counterpart to Fig. I.4 of the main manuscript.

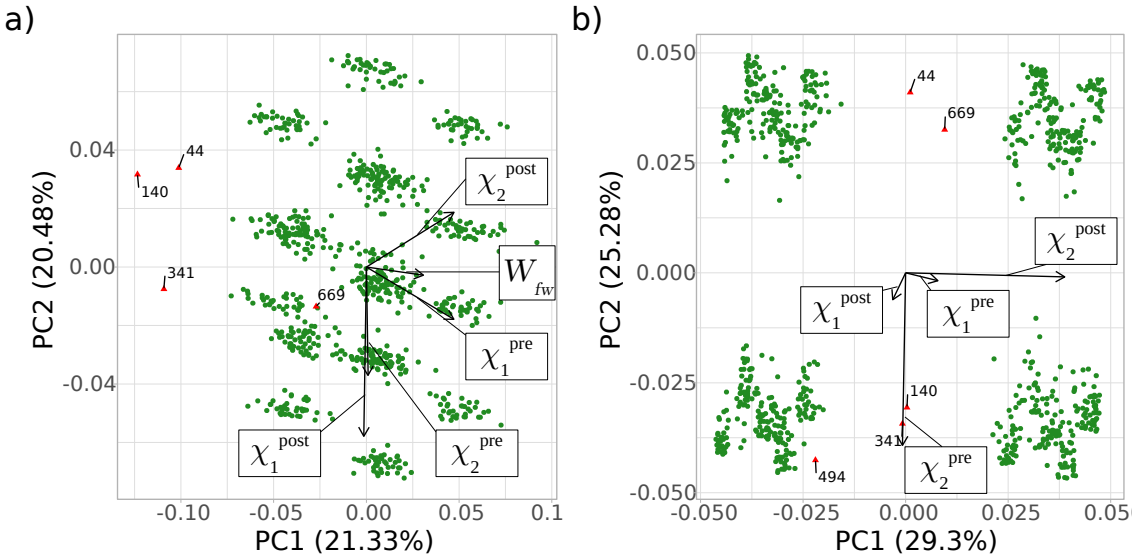


Fig. A.9. PCA of compound **2** plotted as PC1 vs. PC2 for the original protocol. a) Scaled version; b) unscaled version. Outliers are marked according to predefined criteria described in Sect. I.3.3 of the main manuscript; red triangles (rather than green dots) indicate deviations a) lower than $\bar{W} - 3\sigma$, and b) $|\bar{\chi}_2^{post}| - 3\sigma$.

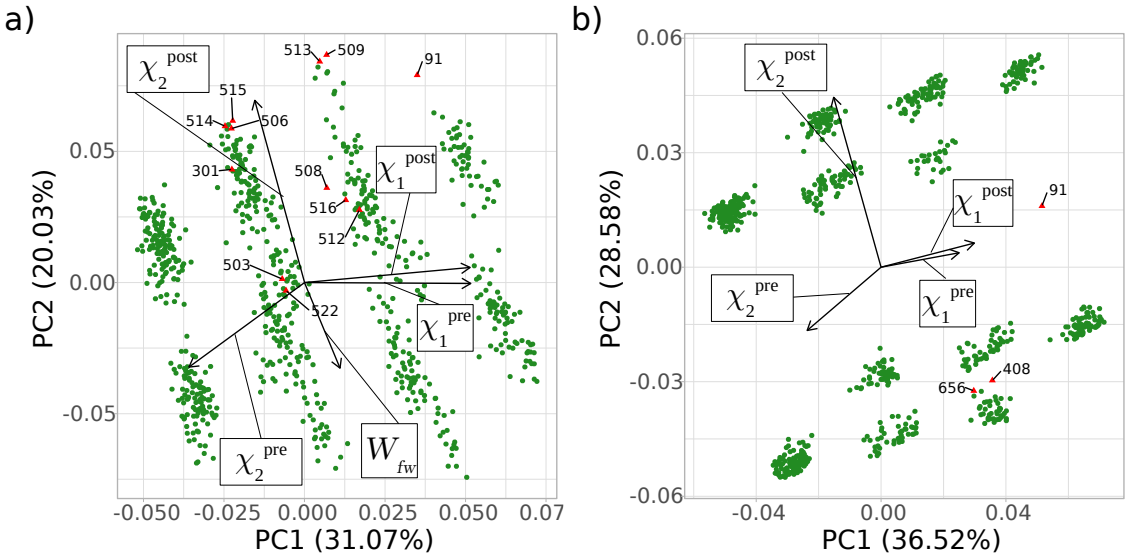


Fig. A.10. PCA of compound **7** plotted as PC1 vs. PC2 for the original protocol. a) Scaled version; b) unscaled version. Outliers are marked according to predefined criteria described in Sect. I.3.3 of the main manuscript; red triangles (rather than green dots) indicate deviations a) lower than $\bar{W} - 3\sigma$, and b) $|\bar{\chi}_2^{post}| - 3\sigma$.

B. Supplementary Material of Publication II

Exploring Routes to Enhance the Calculation of Free Energy Differences via Nonequilibrium Work SQM/MM Switching Simulations Using Hybrid Charge Intermediates between MM and SQM Levels of Theory or Non-Linear Switching Schemes

Tab. B.1. Compound label, total number of atoms (N_{total}), total number of heavy atoms (N_{heavy}), number of TIP3P waters (N_{Waters}), box length and free energy offset for each compound investigated. This offset has to be subtracted (added) to the respective $\Delta A^{MM \rightarrow SQM}$ ($\Delta A^{SQM \rightarrow MM}$) listed in Tables B.3-B.11 below.

Compound	N_{tot}/N_{heavy}	N_{Waters}	Box length [Å]	Offset[kcal/mol]
1-t1	15/7	560	25.588	11000
1-t2	15/7	561	25.802	11000
2-t1	18/11	559	25.630	14000
2-t2	18/11	559	25.576	14000
3-t1	18/11	559	25.595	14000
3-t2	18/11	559	25.596	14000
4-t1	12/7	561	25.589	10000
4-t2	12/7	561	25.573	10000
5-t1	18/11	560	25.622	14000
5-t2	18/11	559	25.582	14000
6-t1	11/7	560	25.562	10000
6-t2	11/7	561	25.557	10000
7-t1	20/10	560	25.612	14000
7-t2	20/10	560	25.581	14000
8-t1	11/7	562	25.559	10000
8-t2	11/7	561	25.556	10000
Ala	22/10	561	25.601	16000
Ser	23/11	561	25.622	18000

Overview of detailed results tables: Tables B.2–B.11 summarize all raw data used to generate the tables and figures shown in the main manuscript. In this study, the free energy differences $\Delta A^{MM \leftrightarrow SQM}$ obtained by CRO (Crook's equation)[41] with switching lengths of 5 ps were considered as the reference results and are referred to as CRO_{Ref}. The deviations $\delta\Delta A$ from CRO_{Ref} for each of the methods studied (cf. Eq. 13 of the main manuscript) are listed in Table B.2. Figures 4 and 5 of the main manuscript are based on the values in Table B.2. Table B.2 also provides the standard error, obtained by Gaussian error propagation of the individual results.

Tables B.3–B.6, B.8 and B.10 contain the detailed results obtained with Jarzynski's equation (JAR)[40] in the forward (MM→SQM/MM) and backward (SQM/MM →MM) direction, as well as the corresponding CRO result using the *same* switching length as for the JAR calculations. In addition to various convergence criteria, see below, we also report the difference between the JAR(fw) and the CRO result with this switching length (rightmost column in Tables B.3 - B.6, B.8 and B.10). Since in Table B.4 we report results obtained with 2 ps switching length, the difference JAR(fw) − CRO reported there is slightly different from the corresponding $\delta\Delta A$ entry in Table B.2 (leftmost column). By contrast, JAR(fw) − CRO in Table B.5, $N_{switch} = 5000$ fs, is identical to the corresponding $\delta\Delta A$ entry in Table B.2. Tables B.6, B.8 and B.10 summarize JAR and CRO results for the three sets of simulations using an intermediate MM state with modified partial charges. To obtain the full free energy difference $\Delta A^{MM \leftrightarrow SQM}$, a correction step, the free energy difference between the force field description of the solute and the description with modified charges, needs to be computed; these results are summarized in Tables B.7, B.9 and B.11. The overall deviation $\delta\Delta A$ of this two-step process from CRO_{Ref} are also reported in the three right-most columns in Table B.2.

Convergence criteria reported in Tables B.3–B.6, B.8 and B.10: In line with earlier work[34], we computed several convergence criteria for each free energy difference $\Delta A^{MM \rightarrow SQM}$. As described in the main manuscript, raw data for each ΔA was generated using eight independent series of calculations. We consider each set of raw data as a block and calculated ΔA_i for each of these blocks. Thus, we can compute a "hysteresis" Hyst(eresis) $= \Delta A - 1/8 \sum_{i=1}^{8} \Delta A_i$; cf. Ref. [75]. Further, we report the one-sided Π values for the forward and backward free energy differences.[76] Π values > 0.5 indicate that the W distribution is based on sufficient and unbiased sampling. We report two variants for Π, one evaluated with the initial guess for ΔA calculated by CRO (Π_{CRO}),the other by JAR (Π_{JAR}).

Additional convergence criteria reported in Tables B.7, B.9 and B.11: In the present work three intermediate MM states were used, which differed from the force field description by the partial charges used for the solute. To complete the thermodynamic cycle, one has to compute the respective free energy difference between the force field (MM) state and the Mulliken charge intermediate state (MULL). Here, this correction $\Delta A^{MM \leftrightarrow MULL}$ was calculated by (two-sided) free energy perturbation, i.e., Bennett's acceptance ratio method (BAR, cf. Sect. *Hybrid charge intermediates* of the main manuscript). The standard deviation $\sigma\Delta A$ was estimated by block averaging. In addition, we report the error for closing the cycle

shown in Fig. 1b of the main manuscript, i.e., the deviation from ideality,

$$\Delta A^{MM \leftrightarrow MULL} + \Delta A^{MULL \rightarrow SQM} - \Delta A^{MM \leftrightarrow SQM} = 0,$$

using the 5 ps CRO results as the reference value. This cycle closure is labelled as BAR+JAR(fw)-CRO$_{Ref}$ in the tables.

Tab. B.2. Comparison of direct switches of 2 and 5 ps length, 2 and 5 ps CRO results, and the three approaches with an intermediate MULL state including standard error. These are the raw data for Figs. 4 and 5 of the main manuscript.

Compound	$\delta\Delta A_{solv}^{MM\rightarrow SQM}$	$\delta\Delta A_{solv}^{MM\leftrightarrow SQM}$	$\delta\Delta A_{solv}^{MM\rightarrow SQM}$	$\delta\Delta A_{solv}^{MM\leftrightarrow MULL(gas)\rightarrow SQM}$	$\delta\Delta A_{solv}^{MM\leftrightarrow MULL(solv)\rightarrow SQM}$	$\delta\Delta A_{solv}^{MM\leftrightarrow MULL(solv*)\rightarrow SQM}$
	JAR(fw)-CRO$_{Ref}$	JAR(fw)-CRO$_{Ref}$	CRO-CRO$_{Ref}$	BAR+JAR(fw)-CRO$_{Ref}$	BAR+JAR(fw)-CRO$_{Ref}$	BAR+JAR(fw)-CRO$_{Ref}$
N_{switch}	2000 fs	5000 fs	2000 fs	2000 fs	2000 fs	2000 fs
$N_{replicate}$	200	200	200	200	200	200
1-t1	0.35 ± 0.12	0.10 ± 0.08	-0.04 ± 0.05	0.21 ± 0.21	-0.13 ± 0.25	-0.06 ± 0.04
1-t2	0.01 ± 0.04	0.01 ± 0.04	0.02 ± 0.03	-0.01 ± 0.11	0.01 ± 0.07	0.01 ± 0.04
2-t1	-0.13 ± 0.13	-0.15 ± 0.12	0.14 ± 0.03	0.12 ± 0.08	0.01 ± 0.09	0.04 ± 0.04
2-t2	-0.10 ± 0.14	-0.05 ± 0.10	0.01 ± 0.05	-0.18 ± 0.35	0.03 ± 0.13	-0.07 ± 0.05
3-t1	0.22 ± 0.04	-0.11 ± 0.15	0.16 ± 0.04	0.01 ± 0.08	-0.02 ± 0.07	0.01 ± 0.04
3-t2	0.14 ± 0.18	0.06 ± 0.13	0.05 ± 0.07	0.09 ± 0.20	0.12 ± 0.15	-0.10 ± 0.07
4-t1	0.02 ± 0.07	-0.13 ± 0.11	0.02 ± 0.03	-0.09 ± 0.07	-0.07 ± 0.10	-0.20 ± 0.04
4-t2	0.08 ± 0.18	0.01 ± 0.10	-0.13 ± 0.06	0.05 ± 0.29	0.06 ± 0.12	-0.04 ± 0.06
5-t1	0.10 ± 0.05	-0.02 ± 0.03	0.03 ± 0.03	-0.02 ± 0.07	0.05 ± 0.07	0.01 ± 0.04
5-t2	1.16 ± 0.15	-0.02 ± 0.14	0.23 ± 0.07	0.32 ± 0.17	0.17 ± 0.17	0.02 ± 0.08
6-t1	0.04 ± 0.08	0.06 ± 0.06	-0.02 ± 0.04	-0.09 ± 0.09	-0.09 ± 0.06	-0.15 ± 0.04
6-t2	1.80 ± 0.42	0.76 ± 0.12	0.04 ± 0.16	0.20 ± 0.20	0.08 ± 0.19	-0.12 ± 0.11
7-t1	-0.09 ± 0.14	0.03 ± 0.06	0.01 ± 0.04	-0.08 ± 0.07	-0.04 ± 0.08	0.02 ± 0.05
7-t2	-0.22 ± 0.23	-0.22 ± 0.17	-0.22 ± 0.06	0.01 ± 0.37	-0.26 ± 0.29	-0.09 ± 0.06
8-t1	0.12 ± 0.05	0.04 ± 0.05	0.11 ± 0.04	-0.07 ± 0.11	-0.06 ± 0.08	-0.06 ± 0.04
8-t2	-0.33 ± 0.12	0.18 ± 0.05	-0.11 ± 0.05	0.08 ± 0.16	-0.09 ± 0.12	-0.11 ± 0.05
Ala	-0.14 ± 0.33	-0.04 ± 0.12	0.02 ± 0.06	-0.49 ± 0.37	-0.03 ± 0.26	-0.10 ± 0.12
Ser	0.13 ± 0.15	0.12 ± 0.12	-0.13 ± 0.07	0.03 ± 0.40	-0.26 ± 0.25	-0.01 ± 0.08
MAD	0.29 ± 0.15	0.12 ± 0.10	0.08 ± 0.05	0.12 ± 0.19	0.09 ± 0.14	0.07 ± 0.06

Tab. B.3. $\Delta A_{gas}^{MM \rightarrow SQM}$ calculated with JAR (fw), JAR (bw), and CRO, as well as relevant convergence metrics, i.e., standard deviation $\sigma \Delta Ai$ for the 8 blocks, the average work $\overline{W}$, the standard deviation of W (σW), the one-sided Π_{CRO} and Π_{JAR} values, as well as percentage overlap for the distribution of work values between the JAR(fw) and JAR(bw) calculations.

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta \Delta A$
1-t1	-356.30	0.01	0.05	-356.19	0.31	0.13	2.46	356.19	0.01	0.07	358.92	2.99	4.11	0.28	-356.64	0.04	26.95	0.35
1-t2	-348.16	0.01	0.03	-348.10	0.25	2.41	2.61	348.14	0.01	0.02	348.22	0.33	2.84	2.54	-348.16	0.03	75.87	-0.01
2-t1	-873.45	0.01	0.03	-873.42	0.19	2.93	2.74	873.41	0.01	0.02	873.44	0.19	2.79	2.73	-873.43	0.03	90.10	-0.02
2-t2	-886.41	0.01	0.02	-886.39	0.15	3.40	2.79	886.37	0.01	0.02	886.39	0.15	2.67	2.80	-886.39	0.03	94.57	-0.02
3-t1	-870.41	0.01	0.02	-870.38	0.19	0.47	2.74	870.47	0.50	1.62	871.31	2.41	7.27	1.36	-870.48	0.03	82.21	0.07
3-t2	-867.93	0.01	0.02	-867.91	0.16	2.88	2.78	867.90	0.01	0.01	867.92	0.16	2.87	2.79	-867.92	0.03	87.11	-0.01
4-t1	-92.83	0.01	0.02	-92.81	0.18	2.77	2.32	92.79	0.01	0.29	92.82	0.19	2.19	2.29	-92.81	0.06	86.93	-0.02
4-t2	-105.54	0.01	0.03	-105.52	0.13	2.61	2.40	105.45	0.01	0.01	105.47	1.13	2.61	2.39	-105.50	0.06	61.56	-0.04
5-t1	-884.59	0.01	0.03	-884.57	0.18	2.42	2.32	884.57	0.01	0.03	884.60	0.17	2.38	2.33	-884.58	0.06	83.05	-0.01
5-t2	-872.67	0.01	0.05	-872.65	0.16	2.42	2.35	872.65	0.01	0.02	872.67	0.16	2.45	2.35	-872.66	0.06	84.63	-0.01
6-t1	-348.91	0.01	0.03	-348.87	0.24	2.73	2.67	348.92	0.01	0.02	348.96	0.21	2.58	2.68	-348.92	0.03	73.11	0.01
6-t2	-330.90	0.01	0.02	-330.88	0.13	2.87	2.83	330.89	0.01	0.01	330.90	0.13	2.91	2.84	-330.89	0.03	87.54	-0.01
7-t1	-737.05	0.01	0.04	-736.97	0.30	2.49	2.54	737.04	0.01	0.04	737.12	0.32	2.59	2.52	-737.04	0.03	64.73	-0.01
7-t2	-684.36	0.01	0.11	-683.67	1.10	2.21	1.53	684.41	0.01	0.06	684.79	0.60	1.27	1.91	-684.34	0.04	36.21	-0.02
8-t1	-203.44	0.01	0.06	-203.41	0.19	2.42	2.30	203.42	0.01	0.04	203.46	0.20	2.25	2.28	-203.43	0.06	80.87	-0.01
8-t2	-224.69	0.01	0.05	-224.67	0.17	2.32	2.33	224.67	0.01	0.03	224.69	0.17	2.48	2.33	-224.68	0.06	85.77	-0.01

Tab. B.4. Detailed results for N_{switch}=2000 fs $N_{replicate}$=200 $\Delta A_{solv}^{MM \to SQM}$

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta\Delta A$
1-t1	-357.92	0.07	0.33	-356.85	1.15	0.17	0.72	358.22	0.44	0.62	360.07	1.53	0.44	0.12	-358.31	0.12	15.61	0.39
1-t2	-348.34	0.01	0.10	-348.14	0.46	1.67	1.79	348.29	0.01	0.08	348.51	0.55	1.99	1.75	-348.31	0.07	55.96	-0.03
2-t1	-872.77	0.11	0.36	-872.32	0.57	1.65	1.38	872.50	0.01	0.13	872.71	0.51	1.97	1.77	-872.50	0.07	50.12	-0.27
2-t2	-891.36	0.12	0.39	-889.99	1.26	0.56	0.46	891.44	0.10	0.34	892.52	1.21	0.54	0.71	-891.27	0.12	20.81	-0.10
3-t1	-868.61	0.01	0.09	-868.37	0.53	1.07	1.72	868.59	0.10	0.43	869.15	1.03	2.02	1.24	-868.67	0.07	52.09	0.06
3-t2	-877.20	0.20	0.51	-875.19	1.64	-0.05	0.02	877.43	0.29	0.59	879.40	1.56	-0.04	0.04	-877.28	0.15	13.68	0.09
4-t1	-91.15	0.03	0.18	-90.90	0.50	1.14	1.69	91.08	0.09	0.39	91.57	0.96	2.11	1.33	-91.16	0.07	50.99	0.01
4-t2	-111.14	0.21	0.51	-109.69	1.42	0.34	0.41	111.22	0.37	0.59	112.91	1.34	0.25	0.24	-111.35	0.13	15.67	0.21
5-t1	-883.22	0.01	0.12	-883.03	0.48	1.48	1.82	883.33	0.01	0.09	883.60	0.66	1.80	1.66	-883.29	0.07	43.44	0.08
5-t2	-863.34	0.11	0.40	-861.67	1.73	-0.52	0.25	864.59	0.22	0.54	867.27	1.97	-0.37	-0.38	-864.27	0.17	10.95	0.93
6-t1	-345.60	0.03	0.21	-345.21	0.67	1.17	1.48	345.69	0.03	0.22	346.18	0.90	1.53	1.34	-345.65	0.08	39.68	0.05
6-t2	-313.78	0.74	1.16	-310.33	2.22	-1.69	-0.79	315.84	2.20	1.60	321.28	2.63	-1.65	-1.66	-315.55	0.43	2.00	1.76
7-t1	-727.52	0.13	0.39	-726.65	0.96	0.93	0.91	727.42	0.02	0.15	728.24	1.03	1.04	0.96	-727.42	0.09	27.88	-0.10
7-t2	-689.96	0.39	0.63	-687.80	1.94	0.31	-0.08	690.51	0.18	0.52	691.60	1.28	-0.07	0.70	-689.96	0.13	16.21	-0.01
8-t1	-200.32	0.01	0.13	-199.98	0.62	1.45	1.55	200.26	0.06	0.26	200.70	0.71	1.59	1.40	-200.33	0.08	44.61	0.01
8-t2	-226.31	0.08	0.33	-225.20	1.14	0.76	0.69	225.98	0.06	0.31	227.08	1.24	0.93	0.69	-226.10	0.10	29.24	-0.21
Ala	-456.14	0.45	0.92	-454.16	1.34	0.06	0.03	456.57	0.12	0.42	457.91	1.41	0.15	0.49	-455.98	0.15	10.98	-0.16
Ser	-534.85	0.13	0.41	-533.06	1.55	-0.17	0.16	535.46	0.14	0.47	537.44	1.89	-0.01	0.04	-535.10	0.16	15.06	0.26

Tab. B.5. Detailed results for N_{switch}=5000 fs $N_{replicate}$=200 $\Delta A_{solv}^{MM \to SQM}$

	JAR(fw)							JAR(bw)							CRO$_Ref$		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta\Delta A$
1-t1	-358.18	0.03	0.22	-357.54	0.91	0.91	1.15	358.18	0.03	0.22	359.10	1.12	1.10	0.86	-358.28	0.09	31.51	0.10
1-t2	-348.34	0.01	0.08	-348.24	0.35	1.90	2.02	348.32	0.01	0.08	348.45	0.41	2.13	1.96	-348.34	0.07	68.09	-0.01
2-t1	-872.79	0.10	0.33	-872.46	0.45	1.15	1.56	872.61	0.09	0.44	872.96	0.83	2.51	1.53	-872.64	0.07	55.19	-0.15
2-t2	-891.31	0.05	0.26	-890.67	0.88	1.14	1.14	891.04	0.10	0.35	891.89	0.95	1.22	0.92	-891.27	0.09	36.60	-0.05
3-t1	-868.94	0.14	0.41	-868.51	0.49	0.86	1.41	868.88	0.25	0.79	869.44	1.09	2.17	1.24	-868.83	0.07	40.55	-0.11
3-t2	-877.28	0.09	0.34	-876.02	1.24	0.53	0.56	877.62	0.14	0.39	878.63	1.17	0.51	0.77	-877.33	0.12	18.59	0.06
4-t1	-91.30	0.07	0.30	-91.02	0.45	1.22	1.64	91.17	0.13	0.57	91.46	0.78	2.53	1.64	-91.18	0.07	55.76	-0.13
4-t2	-111.22	0.05	0.25	-110.34	1.08	0.91	0.89	111.28	0.05	0.27	112.09	1.03	0.88	0.97	-111.22	0.10	28.54	-0.01
5-t1	-883.34	0.01	0.07	-883.24	0.35	2.16	2.04	883.28	0.01	0.08	883.40	0.36	2.06	2.00	-883.32	0.06	68.08	-0.02
5-t2	-864.52	0.10	0.38	-863.11	1.34	0.36	0.44	864.67	0.13	0.47	865.99	1.41	0.47	0.51	-864.50	0.12	20.22	-0.02
6-t1	-345.58	0.02	0.14	-345.38	0.44	1.32	1.79	345.67	0.05	0.29	345.98	0.75	1.96	1.58	-345.63	0.07	42.08	0.06
6-t2	-314.82	0.31	0.61	-312.70	1.75	-0.58	-0.05	316.15	0.35	0.66	318.66	1.91	-0.51	-0.29	-315.59	0.18	8.47	0.76
7-t1	-727.40	0.02	0.16	-727.02	0.68	1.36	1.49	727.43	0.02	0.15	727.87	0.76	1.48	1.40	-727.43	0.08	41.75	0.03
7-t2	-689.96	0.18	0.48	-688.79	1.27	0.97	0.64	689.70	0.06	0.28	690.60	1.05	0.79	0.87	-689.74	0.10	29.56	-0.22
8-t1	-200.39	0.01	0.12	-200.21	0.45	1.21	1.83	200.47	0.03	0.21	200.79	0.90	2.19	1.58	-200.43	0.07	51.75	0.04
8-t2	-225.80	0.01	0.13	-225.46	0.66	1.14	1.55	226.08	0.02	0.18	226.57	0.83	1.37	1.33	-225.98	0.08	34.11	0.18
Ala	-456.03	0.09	0.33	-455.03	1.11	0.77	0.78	456.03	0.07	0.31	457.00	1.12	0.83	0.81	-456.00	0.10	24.98	-0.04
Ser	-534.85	0.06	0.31	-533.83	1.13	0.42	0.76	535.11	0.09	0.35	536.35	1.45	0.72	0.57	-534.97	0.11	22.77	0.12

Tab. B.6. Detailed results for N_{switch}=2000 fs $N_{replicate}$=200 $\Delta A_{solv}^{MULL(gas)\rightarrow SQM}$

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta\Delta A$
1-t1	-300.97	0.09	0.34	-300.10	0.98	0.56	0.90	301.11	0.07	0.30	302.21	1.31	0.94	0.70	-301.03	0.10	26.00	0.07
1-t2	-325.98	0.02	0.18	-325.59	0.67	1.31	1.46	325.95	0.02	0.15	326.42	0.80	1.57	1.35	-325.98	0.08	43.10	-0.01
2-t1	-881.09	0.01	0.11	-880.96	0.39	1.60	1.93	881.24	0.05	0.28	881.45	0.52	1.82	1.77	-881.19	0.07	45.15	0.10
2-t2	-890.10	0.33	0.58	-888.55	1.25	0.43	0.34	889.83	0.11	0.40	891.33	1.36	0.48	0.37	-889.92	0.12	18.51	-0.18
3-t1	-874.39	0.01	0.12	-874.26	0.37	1.89	1.96	874.35	0.06	0.29	874.50	0.42	2.06	1.91	-874.38	0.07	55.87	-0.01
3-t2	-876.09	0.09	0.32	-874.77	1.27	0.37	0.51	875.97	0.14	0.47	877.63	1.43	0.48	0.25	-876.14	0.12	20.59	0.06
4-t1	-96.40	0.06	0.09	-96.28	0.36	1.88	1.96	96.35	0.05	0.30	96.48	0.41	2.19	1.96	-96.37	0.07	62.03	-0.03
4-t2	-109.36	0.16	0.47	-107.99	1.20	0.39	0.47	109.61	0.11	0.37	110.81	1.24	0.49	0.61	-109.36	0.12	16.88	0.01
5-t1	-886.239	0.01	0.10	-886.09	0.42	1.94	1.89	886.20	0.02	0.15	886.36	0.42	1.94	1.90	-886.22	0.07	58.46	-0.02
5-t2	-900.09	0.06	0.26	-899.10	1.05	0.40	0.79	900.45	0.10	0.37	901.73	1.33	0.64	0.54	-900.31	0.11	18.13	0.22
6-t1	-317.77	0.01	0.13	-317.61	0.40	1.83	1.89	317.65	0.04	0.24	317.81	0.47	2.29	1.88	-317.70	0.07	69.77	-0.07
6-t2	-317.66	0.07	0.27	-316.66	1.17	0.44	0.78	317.48	0.43	0.69	319.21	1.44	0.66	0.20	-317.85	0.11	22.24	0.18
7-t1	-753.88	0.01	0.09	-753.74	0.41	1.96	1.94	753.80	0.01	0.06	753.96	0.44	2.06	1.89	-753.85	0.07	69.80	-0.03
7-t2	-781.07	0.31	0.62	-778.59	2.16	0.04	-0.27	781.88	0.14	0.41	783.21	1.46	-0.31	0.50	-781.23	0.15	13.16	0.16
8-t1	-259.93	0.02	0.17	-259.78	0.38	2.01	1.90	259.82	0.01	0.06	259.91	0.34	2.29	2.06	-259.84	0.07	66.17	-0.09
8-t2	-268.62	0.04	0.25	-267.77	1.04	0.82	0.93	268.56	0.17	0.45	269.68	1.13	0.83	0.68	-268.72	0.10	23.42	0.10
Ala	-443.24	0.37	062	-441.70	1.20	0.60	0.34	443.12	0.11	0.39	444.09	1.18	0.62	0.81	-442.89	0.11	21.08	-0.35
Ser	-502.65	0.21	0.66	-501.29	1.28	0.30	0.47	502.89	0.10	0.37	504.23	1.50	0.49	0.50	-502.66	0.12	20.05	0.01

Tab. B.7. Detailed results for energy correction term via FEP $\Delta A_{solv}^{MM\leftrightarrow MULL(gas)}$

	BAR		Cycle closure
Compound	ΔA	σ ΔA	BAR+JAR(fw)-CRO$_{Ref}$
1-t1	-57.25	0.02	0.06
1-t2	-22.37	0.02	-0.01
2-t1	8.58	0.02	0.12
2-t2	-1.35	0.02	-0.18
3-t1	5.57	0.02	0.01
3-t2	-1.16	0.01	0.09
4-t1	5.14	0.02	-0.09
4-t2	-1.81	0.01	0.06
5-t1	2.90	0.01	-0.02
5-t2	35.92	0.03	0.32
6-t1	-27.95	0.02	-0.09
6-t2	2.27	0.04	0.20
7-t1	26.37	0.03	-0.08
7-t2	91.33	0.02	0.01
8-t1	59.42	0.02	-0.07
8-t2	42.72	0.02	0.08
Ala	-13.24	0.02	-0.49
Ser	-32.29	0.02	0.03

Tab. B.8. Detailed results for N_{switch}=2000 fs $N_{replicate}$=200 $\Delta A_{solv}^{MULL(solv)\rightarrow SQM}$

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta\Delta A$
1-t1	-296.98	0.19	0.42	-296.29	0.80	0.89	1.10	296.95	0.05	0.29	297.72	1.08	1.27	1.00	-296.93	0.09	32.64	-0.04
1-t2	-290.29	0.01	0.08	-290.16	0.38	1.82	1.95	290.28	0.01	0.11	290.44	0.45	2.03	1.89	-290.29	0.07	58.94	0.01
2-t1	-872.19	0.01	0.13	-871.99	0.43	1.56	1.79	872.16	0.03	0.23	872.41	0.58	2.02	1.71	-872.18	0.07	52.42	-0.02
2-t2	-866.81	0.03	0.19	-866.42	0.67	1.30	1.47	866.77	0.02	0.15	867.31	0.83	1.50	1.28	-866.84	0.08	41.08	0.03
3-t1	-866.92	0.01	0.09	-866.76	0.39	1.58	1.88	866.92	0.10	0.43	867.13	0.56	2.10	1.77	-866.92	0.07	54.39	0.01
3-t2	-852.65	0.03	0.21	-852.07	0.83	0.99	1.22	852.76	0.07	0.28	853.53	0.95	1.11	1.02	-852.77	0.09	29.55	0.13
4-t1	-90.58	0.02	0.16	-90.44	0.35	1.66	1.92	90.52	0.08	0.41	90.66	0.46	2.43	1.95	-90.53	0.06	62.06	-0.05
4-t2	-87.81	0.02	0.17	-87.39	0.68	1.20	1.43	87.86	0.01	0.13	88.40	0.84	1.43	1.26	-87.86	0.08	34.61	0.06
5-t1	-882.67	0.01	0.10	-882.53	0.39	1.83	1.92	882.73	0.01	0.11	882.85	0.38	1.93	2.00	-882.68	0.07	52.09	0.02
5-t2	-877.79	0.05	0.26	-877.25	0.80	0.99	1.26	878.04	0.02	0.17	878.66	0.93	1.17	1.17	-877.92	0.09	30.33	0.13
6-t1	-319.78	0.01	0.07	-319.64	0.36	1.63	1.93	319.71	0.06	0.35	319.86	0.50	2.48	1.89	-319.73	0.06	66.43	-0.05
6-t2	-306.52	0.05	0.25	-305.44	1.23	0.52	0.71	306.26	0.48	0.67	308.01	1.30	0.55	0.19	-306.72	0.11	21.47	0.20
7-t1	-728.78	0.01	0.10	-728.66	0.37	1.95	1.97	728.77	0.01	0.09	728.90	0.40	2.01	1.95	-728.78	0.07	60.67	-0.01
7-t2	-744.16	0.13	0.48	-742.17	2.17	1.12	0.03	743.78	0.51	0.70	745.00	0.91	0.08	0.59	-744.03	0.13	21.91	-0.13
8-t1	-252.61	0.01	0.10	-252.52	0.31	2.05	2.05	252.49	0.01	0.06	252.56	0.29	2.69	2.13	-252.53	0.06	87.34	-0.08
8-t2	-244.81	0.02	0.18	-244.54	0.54	1.56	1.67	244.69	0.01	0.10	245.03	0.68	1.87	1.56	-244.77	0.07	52.40	-0.04
Ala	-422.18	0.13	0.42	-421.59	0.83	1.10	1.20	422.25	0.05	0.25	422.90	0.88	1.16	1.15	-422.23	0.09	29.88	0.05
Ser	-483.98	0.14	0.40	-483.15	0.97	0.77	0.94	483.85	0.04	0.22	484.78	1.26	1.19	0.85	-483.87	0.09	33.84	-0.11

Tab. B.9. Detailed results for energy correction term via FEP $\Delta A_{solv}^{MM\leftrightarrow MULL(solv)}$

	BAR		Cycle closure
Compound	ΔA	σ ΔA	BAR+JAR(fw)-CRO$_{Ref}$
1-t1	-61.43	0.05	-0.13
1-t2	-58.04	0.04	0.02
2-t1	-0.46	0.01	-0.01
2-t2	-24.43	0.03	0.03
3-t1	-1.96	0.02	-0.04
3-t2	-24.52	0.03	0.17
4-t1	-0.68	0.01	-0.08
4-t2	-23.37	0.03	0.05
5-t1	-0.61	0.01	0.04
5-t2	13.41	0.07	0.11
6-t1	-25.96	0.02	-0.11
6-t2	-8.96	0.22	0.11
7-t1	1.30	0.02	-0.05
7-t2	54.15	0.03	-0.26
8-t1	52.12	0.03	-0.06
8-t2	18.71	0.03	-0.12
Ala	-33.84	0.04	-0.03
Ser	-51.25	0.04	-0.26

Tab. B.10. Detailed results for N_{switch}=2000 fs $N_{replicate}$=200 $\Delta A_{solv}^{MULL(solv*)\rightarrow SQM}$

	JAR(fw)							JAR(bw)							CRO		Overlap (%)	JAR(fw)-CRO
Compound	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	Hyst	σ ΔAi	$\overline{W}$	σW	Π_{CRO}	Π_{JAR}	ΔA	σ ΔA	W	$\delta\Delta A$
1-t1	-255.56	0.01	0.05	-255.46	0.33	1.69	2.04	255.59	0.01	0.05	255.76	0.53	2.16	1.86	-255.59	0.06	57.14	0.04
1-t2	-280.90	0.01	0.11	-280.76	0.40	1.69	1.91	280.87	0.01	0.09	281.10	0.55	2.04	1.74	-280.91	0.07	55.25	0.01
2-t1	-871.50	0.01	0.12	-871.36	0.37	2.03	1.94	871.43	0.01	0.08	871.51	0.31	2.21	2.10	-871.43	0.06	63.71	-0.07
2-t2	-845.45	0.01	0.06	-845.40	0.24	2.33	2.22	845.40	0.01	0.04	845.44	0.23	2.35	2.23	-845.42	0.06	86.69	-0.02
3-t1	-867.80	0.01	0.10	-867.70	0.33	1.87	2.03	867.70	0.02	0.18	867.80	0.38	2.57	2.03	-867.74	0.06	81.16	-0.05
3-t2	-821.91	0.01	0.06	-821.85	0.27	2.25	2.16	821.82	0.01	0.05	821.89	0.29	2.46	2.12	-821.87	0.06	88.84	-0.04
4-t1	-90.98	0.02	0.17	-90.83	0.38	1.64	1.88	90.82	0.04	0.27	90.95	0.45	2.82	1.94	-90.87	0.07	74.48	-0.11
4-t2	-65.33	0.01	0.05	-65.29	0.21	2.24	2.25	65.30	0.01	0.04	65.35	0.23	2.40	2.24	-65.32	0.06	80.71	-0.02
5-t1	-879.94	0.01	0.10	-879.84	0.32	1.90	2.02	879.92	0.01	0.08	880.02	0.38	2.23	2.03	-879.92	0.06	63.06	-0.02
5-t2	-870.07	0.01	0.06	-870.01	0.25	2.20	2.18	870.02	0.01	0.04	870.08	0.27	2.34	2.17	-870.05	0.06	80.10	-0.02
6-t1	-315.73	0.01	0.06	-315.57	0.37	1.55	1.88	315.56	0.01	0.11	315.68	0.43	3.12	2.00	-315.61	0.06	77.97	-0.12
6-t2	-279.51	0.01	0.06	-279.48	0.20	2.24	2.27	279.50	0.01	0.04	279.54	0.21	2.36	2.28	-279.50	0.06	77.40	-0.01
7-t1	-739.64	0.01	0.06	-739.57	0.28	2.04	2.14	739.67	0.01	0.07	739.75	0.32	2.08	2.09	-739.66	0.06	61.21	0.02
7-t2	-706.62	0.03	0.22	-704.88	2.05	1.44	0.20	706.81	0.05	0.27	707.35	0.65	0.16	1.27	-706.56	0.12	24.94	-0.06
8-t1	-253.13	0.01	0.05	-253.03	0.31	2.07	2.03	253.03	0.01	0.04	253.08	0.23	2.63	2.22	-253.05	0.06	71.13	-0.08
8-t2	-226.64	0.01	0.05	-226.60	0.22	2.27	2.24	226.61	0.01	0.04	226.65	0.23	2.37	2.24	-226.62	0.06	86.52	-0.02
Ala	-393.52	0.16	0.39	-393.08	0.68	1.25	1.40	393.59	0.02	0.16	394.02	0.78	1.48	1.41	-393.52	0.08	38.17	0.01
Ser	-449.92	0.06	0.29	-449.52	0.70	0.98	1.45	450.01	0.04	0.23	450.69	1.05	1.49	1.11	-450.02	0.08	37.71	0.10

Tab. B.11. Detailed results for energy correction term via FEP $\Delta A_{solv}^{MM\leftrightarrow MULL(solv*)}$

	BAR		Cycle closure
Compound	ΔA	σ ΔA	BAR+JAR(fw)-CRO$_{Ref}$
1-t1	-102.78	0.03	-0.06
1-t2	-67.42	0.02	0.01
2-t1	-1.10	0.01	0.04
2-t2	-45.89	0.03	-0.07
3-t1	-1.02	0.03	0.01
3-t2	-55.53	0.03	-0.11
4-t1	-0.39	0.01	-0.20
4-t2	-45.93	0.03	-0.04
5-t1	-3.38	0.02	0.01
5-t2	5.59	0.03	0.02
6-t1	-30.05	0.02	-0.15
6-t2	-36.20	0.05	-0.12
7-t1	12.22	0.02	0.02
7-t2	16.80	0.02	-0.09
8-t1	52.64	0.02	-0.06
8-t2	0.55	0.02	-0.11
Ala	-62.58	0.03	-0.10
Ser	-85.06	0.03	-0.01

Tab. B.12. Compound, Root Mean Square Deviation of atomic charges relative to SQM for MM-SQM($RMSD_q$[C]$_{MM}$), MULL(gas)-SQM($RMSD_q$[C]$_{Mull(gas)}$) and MULL(solv)-SQM($RMSD_q$[C]$_{Mull(solv)}$). Differential dipole moment MM-SQM($\Delta\mu_{MM}$), differential dipole moment MULL(gas)-SQM($\Delta\mu_{MULL(gas)}$), differential dipole moment MULL(solv)-SQM($\Delta\mu_{MULL(solv)}$), angle between dipole moments relative to SQM for MM, MULL(gas) and MULL(solv))

Compound	$RMSD_q$[e]			$\Delta\mu$[D]			θ_{SQM}[°]		
	MM	$MULL(gas)$	$MULL(solv)$	MM	$MULL(gas)$	$MULL(solv)$	MM	$MULL(gas)$	$MULL(solv)$
1-t1	0.16	0.09	0.06	4.65	2.87	2.61	52.4	17.3	25.7
1-t2	0.08	0.06	0.01	0.15	0.56	0.09	2.2	2.0	1.3
2-t1	0.06	0.02	0.01	1.67	0.57	0.91	20.7	13.8	15.9
2-t2	0.16	0.09	0.05	3.92	5.31	2.72	1.0	0.3	1.7
3-t1	0.07	0.02	0.01	1.45	0.24	0.98	31.4	0.3	23.1
3-t2	0.16	0.09	0.06	5.91	5.63	3.78	18.3	9.4	10.0
4-t1	0.07	0.02	0.01	1.42	0.21	0.80	17.7	8.0	13.4
4-t2	0.20	0.11	0.07	3.72	4.49	2.52	12.5	6.4	7.8
5-t1	0.05	0.02	0.01	1.89	0.28	1.00	37.5	18.7	25.2
5-t2	0.18	0.08	0.06	6.74	4.14	3.21	49.3	9.1	24.3
6-t1	0.13	0.02	0.02	1.04	0.42	0.58	32.3	8.0	19.3
6-t2	0.29	0.11	0.10	8.10	4.17	4.22	58.3	10.6	31.0
7-t1	0.14	0.02	0.02	3.35	0.78	0.99	86.4	13.3	24.4
7-t2	0.17	0.11	0.05	6.96	6.54	3.45	8.1	1.5	4.7
8-t1	0.16	0.02	0.02	1.50	0.36	0.62	85.1	1.3	48.1
8-t2	0.24	0.10	0.05	2.94	3.61	2.00	2.2	12.3	6.5
Ala	0.12	0.07	0.04	2.82	2.11	1.16	5.0	0.4	0.7
Ser	0.13	0.07	0.04	2.61	1.67	1.01	16.7	3.5	4.2
MAD	0.14	0.06	0.04	3.38	2.44	1.81	29.8	7.6	16.0

Tab. B.13. Resulting parameters from mono-exponential fitting of the relaxation process for a set of compounds; ΔU_0 represents the initial value, τ is the resulting time constant of the fit, ΔU_{2000} is the difference in energy to the baseline (≈0 kcal/mol) after 2000 fs of relaxation time.

	MM			MULL(solv)		
Compound	ΔU_0 [kcal/mol]	τ [fs^{-1}]	ΔU_{2000} [kcal/mol]	ΔU_0 [kcal/mol]	τ [fs^{-1}]	ΔU_{2000} [kcal/mol]
4-t2	8.2	996	1.1	3.5	1052	0.5
5-t2	10.6	1075	1.7	4.2	908	0.5
6-t2	19.6	744	1.3	6.3	933	0.7
8-t2	4.3	1022	0.6	2.1	814	0.2

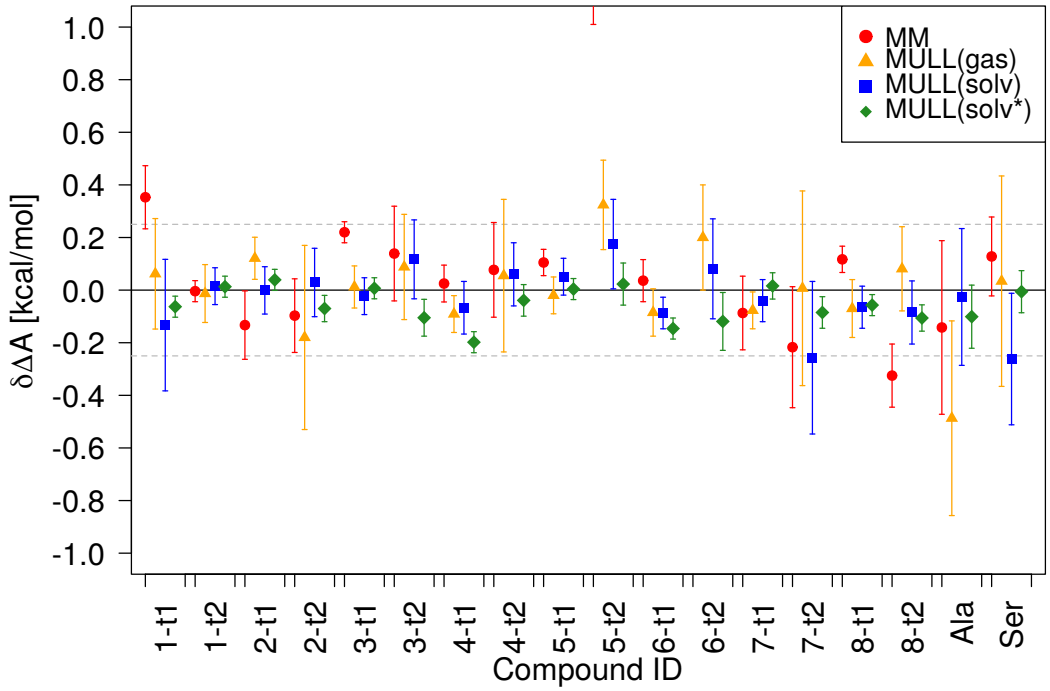


Fig. B.1. $\delta\Delta A$ calculated as the difference between one-sided method JAR(fw) and two-sided method CRO_{Ref} for switching forward($MM^* \rightarrow SQM$) from standard MM charges (MM), Hybrid charge intermediate based on MM gas phase simulation (MULL(gas)), Hybrid charge intermediate based on MM explicit solvent simulation (MULL(solv)) and Hybrid charge intermediate based on SQM/MM explicit solvent simulation (MULL(solv*)) including standard error.

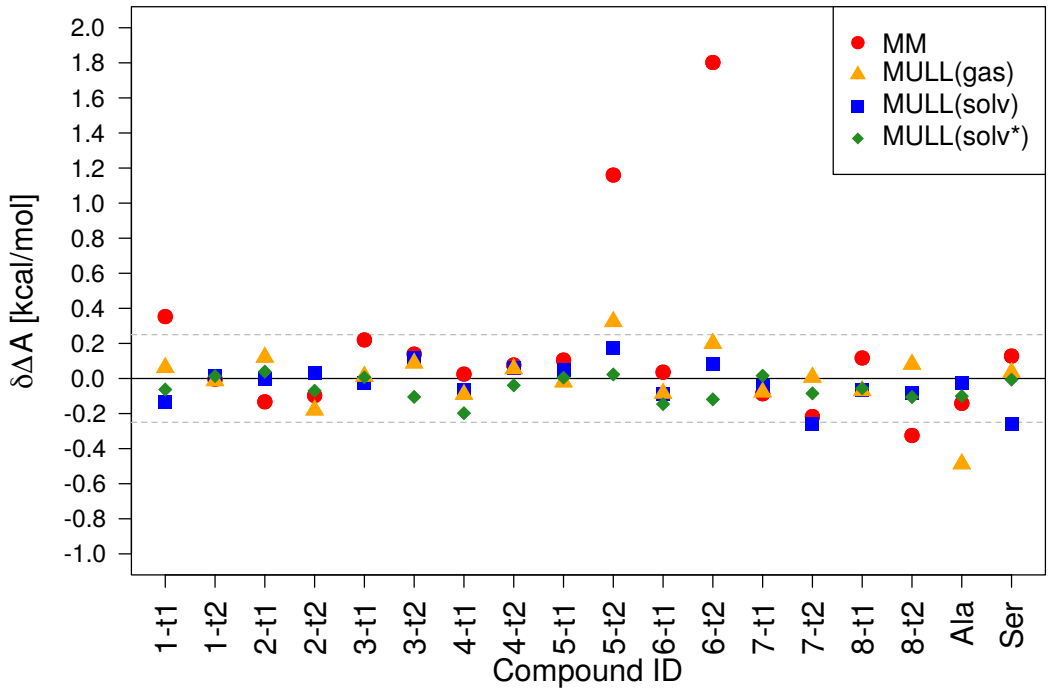


Fig. B.2. $\delta\Delta A$ calculated as the difference between one-sided method JAR(fw) and two-sided method CRO_{Ref} for switching forward($MM^* \rightarrow SQM$) from standard MM charges (MM), Hybrid charge intermediate based on MM gas phase simulation (MULL(gas)), Hybrid charge intermediate based on MM explicit solvent simulation (MULL(solv)) and Hybrid charge intermediate based on SQM/MM explicit solvent simulation (MULL(solv*)).

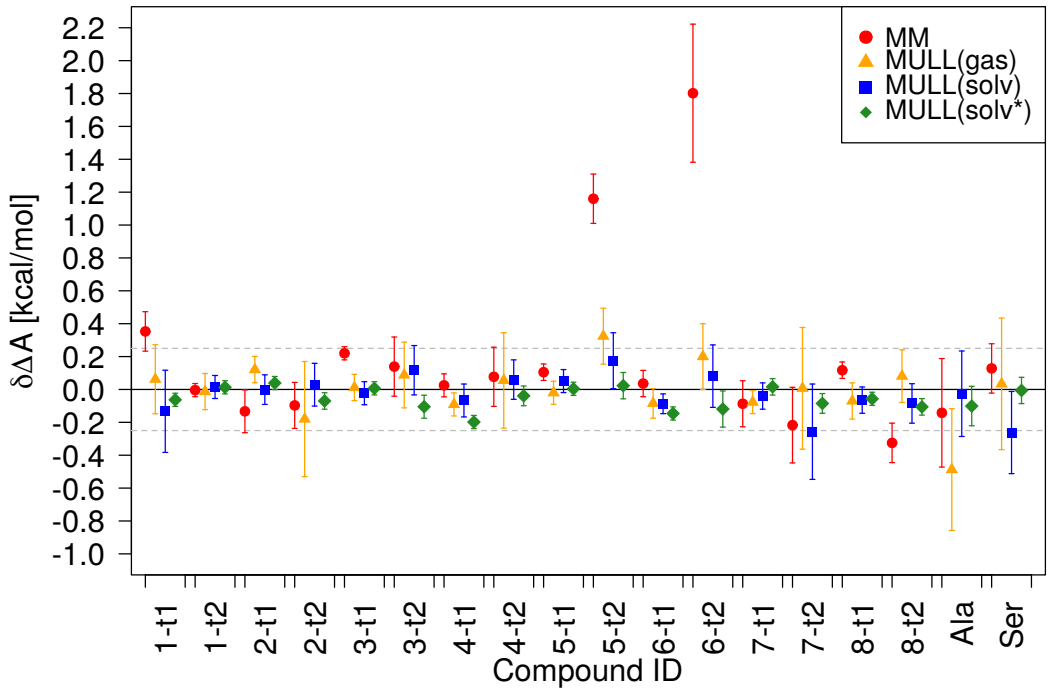


Fig. B.3. $\delta\Delta A$ calculated as the difference between one-sided method JAR(fw) and two-sided method CRO_{Ref} for switching forward($MM^* \rightarrow SQM$) from standard MM charges (MM), Hybrid charge intermediate based on MM gas phase simulation (MULL(gas)), Hybrid charge intermediate based on MM explicit solvent simulation (MULL(solv)) and Hybrid charge intermediate based on SQM/MM explicit solvent simulation (MULL(solv*)) including standard error.

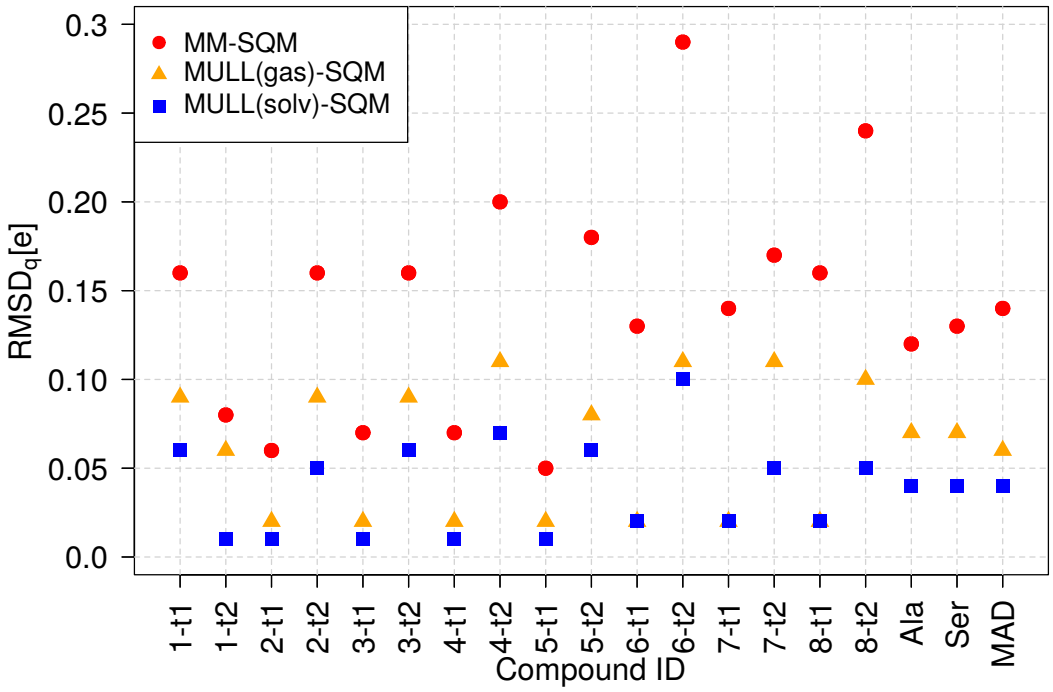


Fig. B.4. Root Mean square deviation of partial charge for all compounds; the MAD is plotted as well.

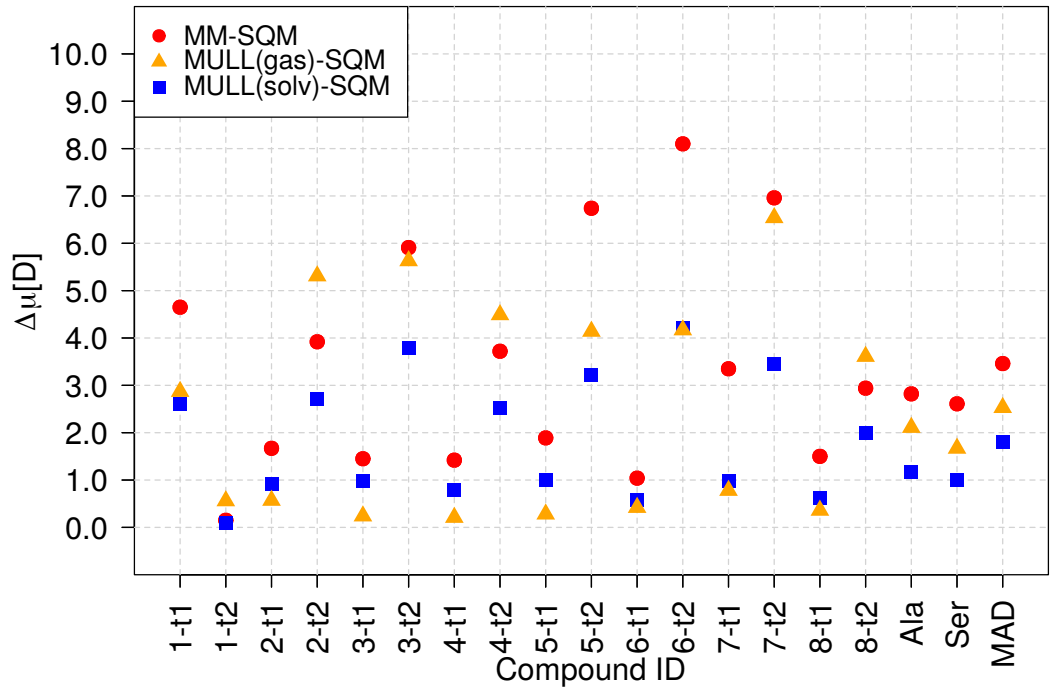


Fig. B.5. Differential dipole moment for all compounds, including MAD.

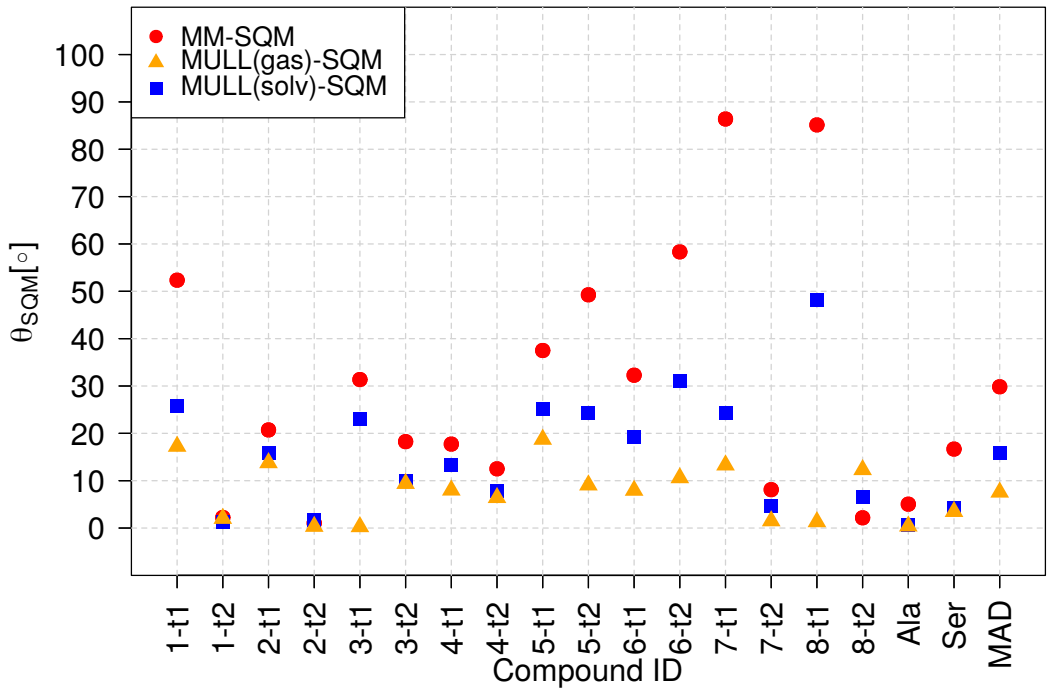


Fig. B.6. Angle between dipole moments for all compounds to SQM dipole moment, including MAD.

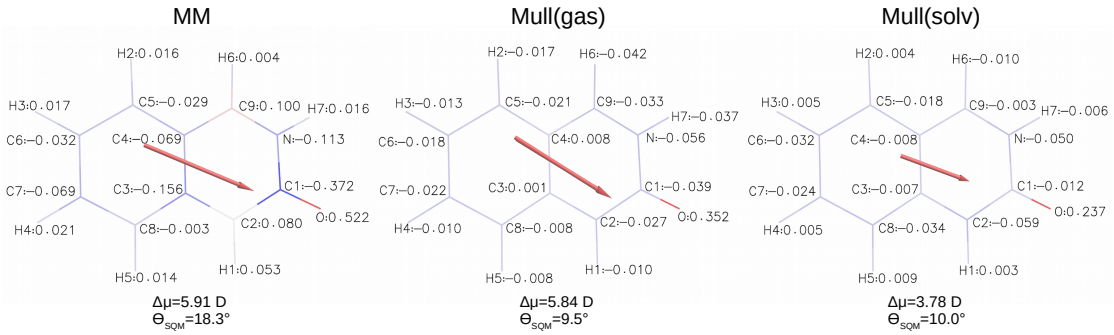


Fig. B.7. Differential dipole moments and atomic charges for **3-t2** including angle to SQM.

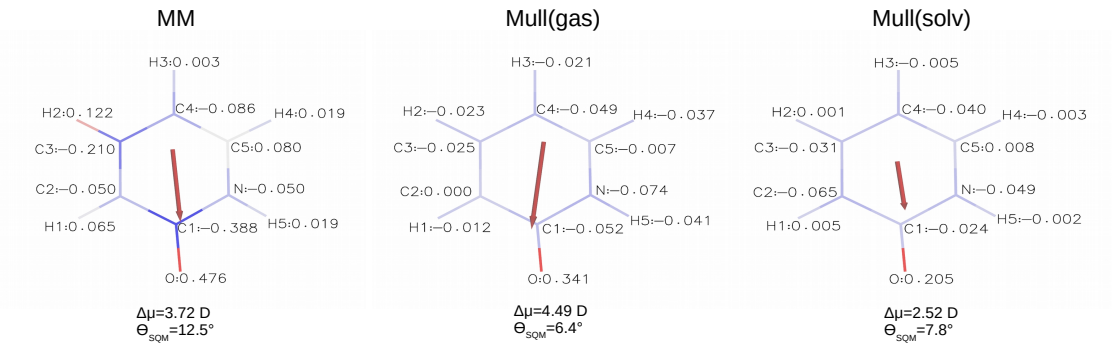


Fig. B.8. Differential dipole moments and atomic charges for **4-t2** including angle to SQM.

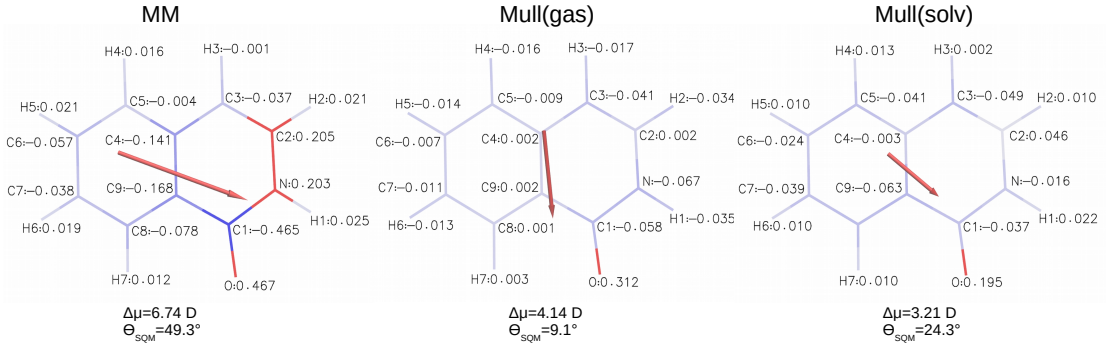


Fig. B.9. Differential dipole moments and atomic charges for **5-t2** including angle to SQM.

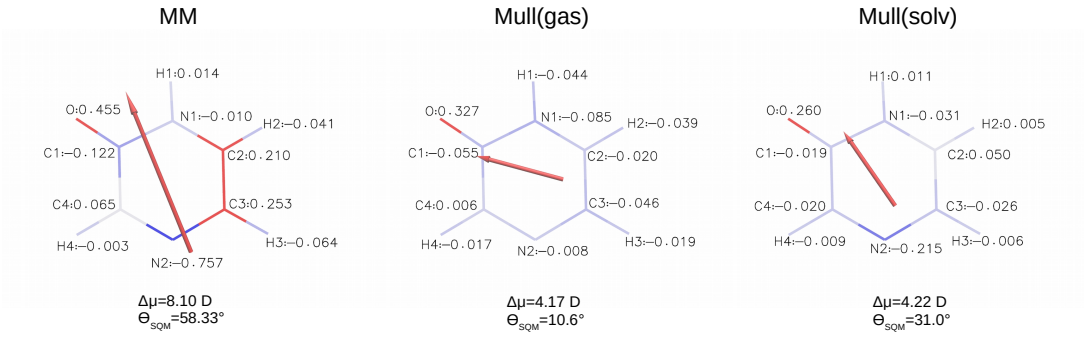


Fig. B.10. Differential dipole moments and atomic charges for **6-t2** including angle to SQM.

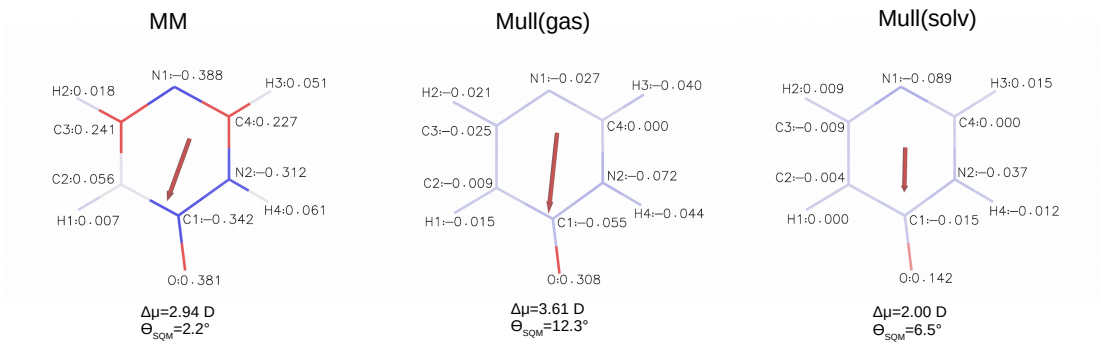


Fig. B.11. Differential dipole moments and atomic charges for **8-t2** including angle to SQM.

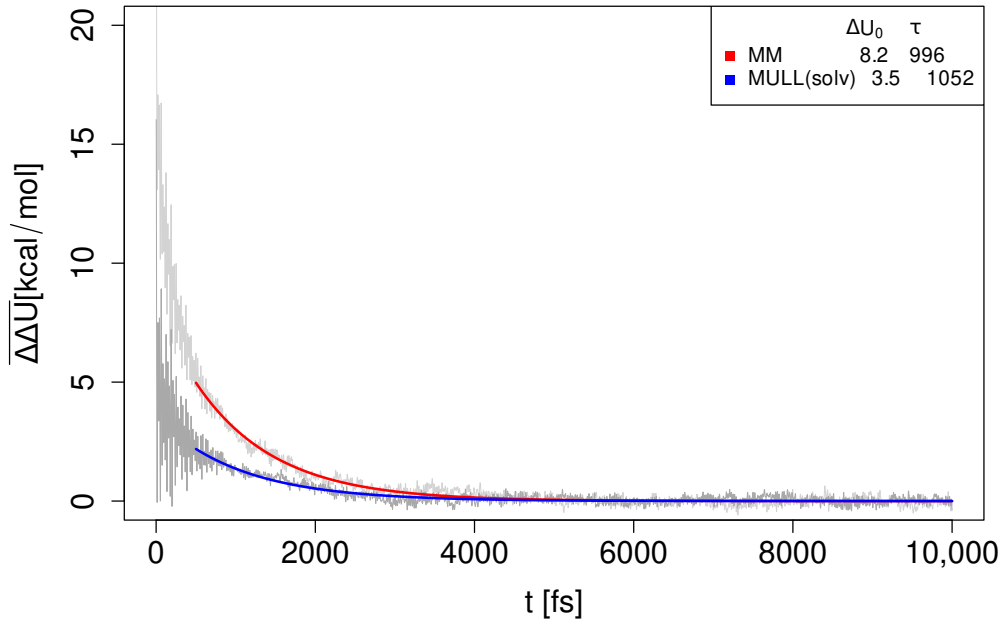


Fig. B.12. Relaxation of TIP3P Water for **4-t2** when switching from MM and MULL(solv) to SQM.

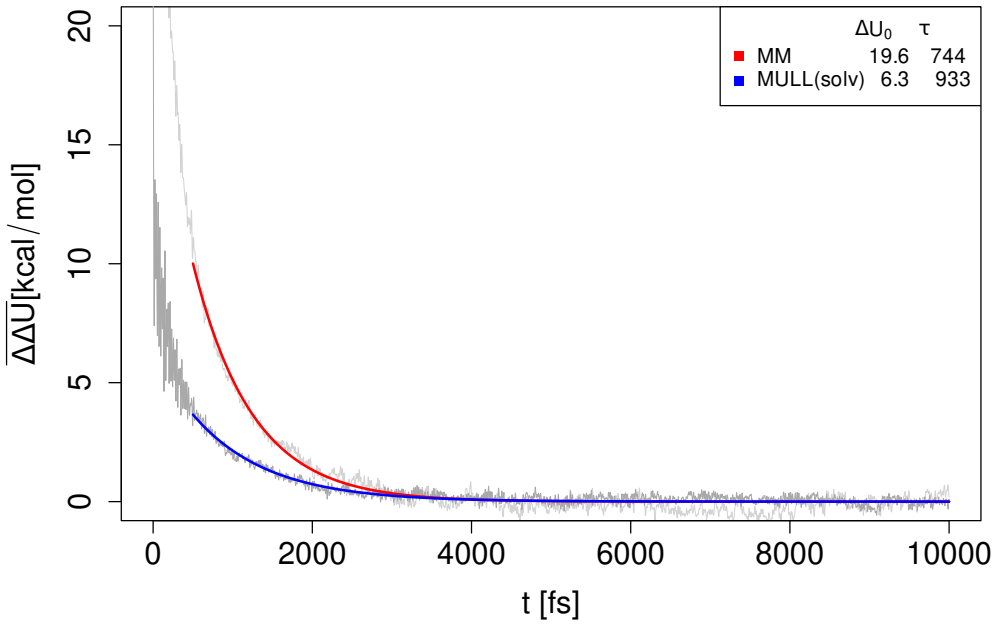


Fig. B.13. Relaxation of TIP3P Water for **6-t2** when switching from MM and MULL(solv) to SQM.

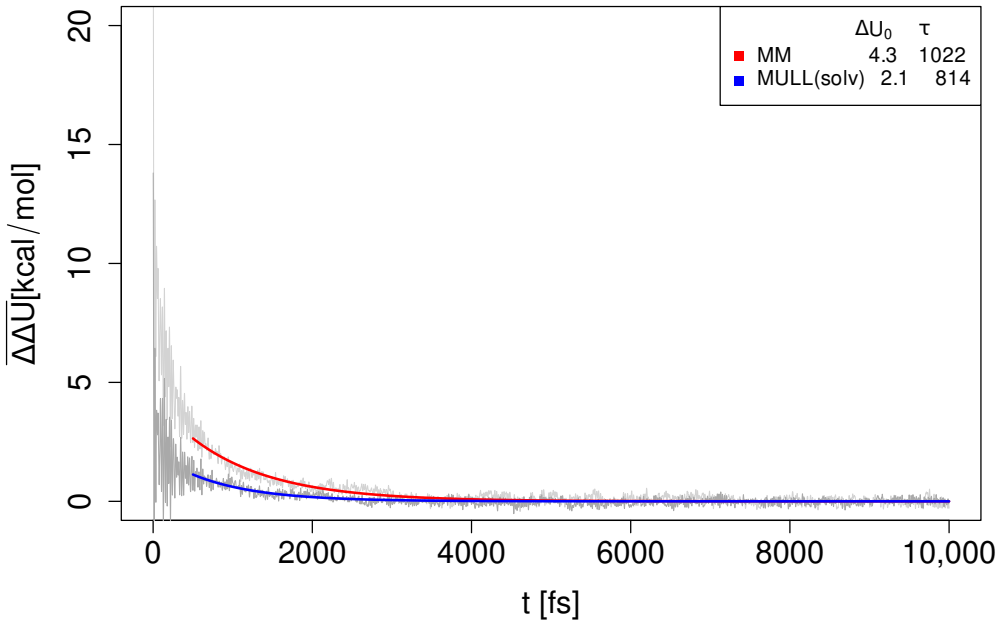


Fig. B.14. Relaxation of TIP3P Water for **8-t2** when switching from MM and MULL(solv) to SQM.

Complete List of Publications

Schöller, A., Kearns, F., Woodcock, H. L., Boresch, S. Optimizing the Calculation of Free Energy Differences in Nonequilibrium Work SQM/MM Switching Simulations *J. Phys. Chem. B* **2022** *126*, 2798–2811. American Chemical Society. DOI: 10.1021/acs.jpcb.2c00696.

Schöller, A. Woodcock, H. L., Boresch, S. Exploring Routes to Enhance the Calculation of Free Energy Differences via Nonequilibrium Work SQM/MM Switching Simulations Using Hybrid Charge Intermediates between MM and SQM Levels of Theory or Non-Linear Switching Schemes *Molecules* **2023** *28*, 4006. MDPI. DOI: 10.3390/molecules28104006.

Tkaczyk, S., Karwounopoulos, J., Schöller, A., Woodcock, H. L., Langer, T., Boresch, S., & Wieder, M. Using Neural Network Potentials to efficiently calculate indirect free energy estimates *ChemRxiv* **2023** DOI:10.26434/chemrxiv-2023-qq206.