Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Quantum Chemical Investigations on the Discrimination of RNA Polymerase II and Systematic Studies of the QM Size Convergence

Sven Roßbach

aus

Nürtingen

2017

Erklärung
Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. C. Ochsenfeld betreut.

Eidesstattliche Versicherung
Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 18.06.2017

(Sven Roßbach)

Dissertation eingereicht am:	25.04.2017
1. Gutachter:	Prof. Dr. Christian Ochsenfeld
2. Gutachter:	Prof. Dr. Regina de Vivie-Riedle
Mündliche Prüfung am:	12.06.2017

Danksagung

Bei Prof. Dr. Christian Ochsenfeld bedanke ich mich für die interessante Themenstellung, die gute Betreuung und stetige Unterstützung.

Prof. Dr. Regina de Vivie-Riedle danke ich für die Zeit und Mühe der Erstellung des Zweitgutachtens.

Meinen Kolleginnen und Kollegen der Arbeitsgruppe danke ich für die angenehme Atmosphäre und hilfreichen Diskussionen – im wissenschaftlichen Rahmen und darüber hinaus.

Besonders danken möchte ich meiner Kollegin und Lebensgefährtin Iris Blank, für ausführliche Diskussionen und ihre tatkräftige Unterstützung. Außerdem danke ich Iris für all das, was neben der Arbeit im Leben noch wichtig ist.

Auch meinen Eltern gilt ein besonderer Dank für ihre kontinuierliche Unterstützung und Begleitung auf meinem Weg durch das Studium.

Zusammenfassung

Im ersten Teil dieser Arbeit wird untersucht wie das Enzym RNA Polymerase II (Pol II) in der Polymerisationsreaktion zwischen Nukleosidtriphosphat (NTP) und 2'-desoxyribo-Nukleosid-triphosphat (dNTP) unterscheidet. In Analogie zu anderen Polymerasen wurden bis zu drei Wasserstoffbrücken zum 2'-Sauerstoff des NTPs postuliert, die für dessen Erkennung zuständig sein können. Ausgehend von einer Kristallstruktur mit allen drei Wasserstoffbrücken zeigen wir, dass in Molekulardynamik Simulationen nur eine Wasserstoffbrücke stabil ist. Von den Mechanismen für die Polymerisationsreaktion, die basierend auf einer anderen Kristallstruktur berechnet und vorgeschlagen wurden, haben wir die vielversprechendsten auf unser Setup angewandt und nur der Mechanismus mit OH^- als Base ist energetisch möglich und liefert eine Reaktionsbarriere mit guter Übereinstimmung zum Experiment. Durch den Vergleich der Reaktionen von NTP und dNTP finden wir, in Übereinstimmung mit experimentellen Mutationsstudien, dass die Aminosäure Arg446 die entscheidende Rolle spielt. Wir zeigen welche Folgen das leicht geänderte Wasserstoffbücken-Netzwerk für die Reaktion hat und wie durch den Verlust dieser einen Wasserstoffbrücke effektiv diskriminiert werden kann. Im Gegensatz zum Experiment erklären wir detailliert warum diese Aminosäure wichtig ist und wie sich die Konformationsänderungen auf das Reaktionsprofil auswirken. Um unsere Ergebnisse zu unterstützen, haben wir die Änderung der Freien Energie durch einen Quanten-Mechanik/Molekular-Mechanik-Freie Energie Störungs-Ansatz (QM/MM-FEP) berücksichtigt und den Einfluss der QM Größe untersucht. 800 QM Atome sind nötig um Reaktionsbarriere und Reaktionsenergie mit der QM Größe zu konvergieren.

Während im Rahmen der vorliegenden Arbeit, für die Untersuchung der Diskriminierung im chemischen Schritt des Mechanismus der Pol II, ein additives QM/MM Schema mit elektrostatischer Einbettung verwendet wurde, wird im zweiten Teil dieser Arbeit am Beispiel eines viel untersuchten Protonentransfers in einem DNA Doppelstrang untersucht, wie Kopplungs- und Einbettungs-Schemata die QM Größenkonvergenz beeinflussen und welche QM Größe für solche Systeme nötig sind. Wir finden, dass das Kopplungs-Schema nur einen geringen Einfluss auf die QM Größenkonvergenz hat, wohingegen das Einbettungs-Schema wichtig ist. Reine QM Berechnungen im Vakuum und mit mechanischer Einbettung zeigen eine erhebliche Abhängigkeit von der QM Größe und konvergieren erst mit 4000-6000 QM Atomen. Für unser System führt elektrostatische Einbettung schnell und mit geringen Schwankungen zu einer Konvergenz bei etwa 1000 QM Atomen.

Das subtraktive ONIOM („*Our own N-layered Integrated molecular Orbital and molecular Mechanics*") Schema von Morokuma braucht im Vergleich mit einem additiven QM/MM Schema mehr einzelne Rechnungen, lässt sich allerdings auf beliebig viele Schichten mit beliebigen Methoden erweitern. Im dritten Teil dieser Arbeit wird der Einfluss einer zweiten QM Schicht in einem QM/QM/MM ONIOM Schema mit elektrostatischer Einbettung auf die vertikalen Anregungsenergien von Methylenblau untersucht. Da für elektronische Anregungsenergien sehr genaue Methoden eingesetzt werden müssen, die auf relativ wenige QM Atome beschränkt sind, kann eine größere, zweite QM Schicht helfen den Einfluss der Umgebung mit einer weniger genauen Methode zu extrapolieren. Wir betrachten zwei unterschiedliche Bindungsmodi von Methylenblau an DNA (kleine Furche, Interkalation) und zerlegen die Änderung durch die zusätzliche Schicht in einzelne Beiträge: Polarisation der inneren QM Schicht

durch eine geänderte Ladungsverteilung der äußeren QM Schicht, elektronische Kopplung der beteiligten Zustände durch Mischen mit elektronischen Zuständen der Umgebung und den Einfluss der Geometrieoptimierung der näheren Umgebung auf QM Niveau. Den größten Einfluss, wenn vorhanden, hat die elektronische Kopplung wie wir am Beispiel der Interkalation zeigen. Die anderen Beiträge sind wesentlich kleiner und heben sich häufig auf, sodass die QM/MM Beschreibung mit nur zwei Schichten ein sehr ähnliches Ergebnis liefert.

Abstract

In the first part of this work, it is investigated how the enzyme RNA Polymerase II (Pol II) distinguishes in the polymerization reaction between nucleoside triphosphate (NTP) and 2'-deoxyribo nucleoside triphosphate (dNTP). By analogy to other polymerases, hydrogen bonds between 2'-oxygen and up to three amino acids were postulated that could be responsible for the recognition of the sugar motif. Based on a crystal structure of Pol II containing all three hydrogen bonds, we show that only one hydrogen bond is stable in molecular dynamics simulations. Out of the mechanisms calculated and suggested based on another crystal structure, we applied the most promising one to our setup and found only the mechanism with OH^- as base to be energetically feasible, leading to a reaction barrier that is in good agreement with the experiment. Comparing the reactions with NTP and dNTP, we find that the amino acid Arg446 is crucial for discrimination, which is in agreement with mutation studies. We demonstrate the consequences of the slightly modified hydrogen bond network on the reaction and how the loss of this single hydrogen bond can lead to an effective discrimination. In contrast to the experiments, we explain in detail why Arg446 is important and how the conformational change affects the reaction profile. To further support our calculations, we took the change in free energy into account by a quantum mechanics / molecular mechanics - free energy perturbation approach (QM/MM-FEP) and studied the influence of the QM size. 800 QM atoms are necessary to properly converge reaction barrier and reaction energy with the QM size.

For the investigation of the discrimination process in the chemical reaction of Pol II, we employed an additive QM/MM scheme with electrostatic embedding. In the second part of this work, we study the influence of different coupling and embedding schemes on the QM size convergence at the example of an extensively studied proton transfer within a DNA double strand and identify the necessary QM size that should be considered by future studies. We find that the coupling scheme has only a small influence on the QM size convergence, whereas the embedding scheme is very important. The results of pure QM calculations in vacuum and calculations employing mechanical embedding show a significant dependency on the QM size and converge only with 4000 to 6000 QM atoms. For our system, electrostatic embedding leads to a fast convergence with small fluctuations, resulting in a converged QM region of about 1000 atoms.

Compared to an additive QM/MM scheme, the subtractive ONIOM („O*ur own* N-*layered* I*ntegrated molecular* O*rbital and molecular* M*echanics*“) scheme of Morokuma requires more calculations, but allows an extension to any number of layers combined with any level of theory. In the third part of this work, the influence of a second QM layer within a three layer QM/QM/MM-ONIOM scheme with electrostatic embedding is investigated on the vertical excitation energies of methylene blue. To calculate electronic excitation energies, very accurate methods are necessary that are limited to relatively few atoms. Therefore, a larger second QM layer can be employed with a faster, less expensive, but also less accurate method to extrapolate the influence of the environment. We study two different binding modes of methylene blue to DNA (minor-groove, intercalation) and decompose the influence of the additional layer into separate contributions: polarization of the inner QM layer by the modified charge distribution of the outer QM layer, electronic-coupling of the involved states due to mixing with electronic states of the environment, and the influence of the geometry optimization of the nearer surroundings on the QM level. For the example of intercalated methylene blue, we show that

electronic-coupling has the most important influence, if present. All other contributions are significantly smaller and often cancel each other, which is the reason for the good performance of the QM/MM scheme with only two layers.

Contents

List of publications

This work is a cumulative dissertation. While two papers have been published already, the third project is summarized in a manuscript that will be submitted soon. These papers are considered the main part of this thesis. The following list specifies the contributions of the author in each of the publications in italics. The publications and their corresponding supplementary information are included in this work and are referred to as paper 1-3.

Paper 1: “Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II, S. Roßbach, C. Ochsenfeld,
J. Chem. Theory Comput., **2017**, *13*, 1699-1705.
Contributions by S. Roßbach: ideas, all calculations, analysis, and writing

Paper 2: "Influence of coupling and embedding schemes on the QM size convergence in QM/MM approaches for the example of a proton transfer in DNA", S. Roßbach, C. Ochsenfeld,
J. Chem. Theory Comput., **2017**, *13*, 1102-1107.
Contributions by S. Roßbach: ideas, all calculations, analysis, and writing

Paper 3: “Effect of DNA Environment on Electronically Excited States of Methylene Blue in a QM/QM/MM Scheme”, J. J. Nogueira, S. Roßbach, C. Ochsenfeld, L. González, *to be submitted.*
Contributions by S. Roßbach: implementation of the ONIOM approach and the interface to Molcas in ChemShell, and partial writing.

Introduction

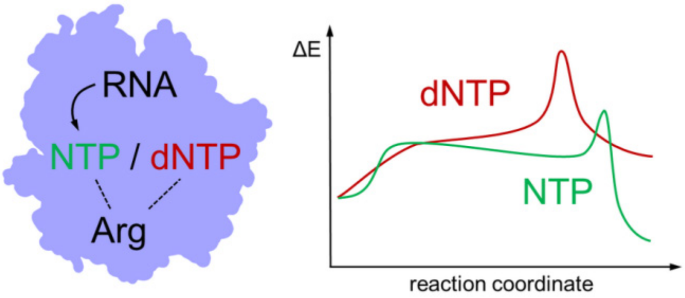


Figure 1. Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II.

The first part of this thesis is about discrimination in the chemical reaction catalyzed by RNA polymerase II. RNA polymerase II is the enzyme in eukaryotic cells that transcribes protein-coding genes (DNA) into a complementary chain of ribonucleotides (RNA), which will eventually be translated to a protein. All cellular RNA polymerases share a highly conserved catalytic core,[1–3] consisting of 10 subunits with additional subunits on the periphery. A general two-ion mechanism is assumed[4,5] in all cases, although details may be different. Structural studies have provided detailed information about the different factors and conformational changes involved in the nucleotide addition cycle,[6,7] which lead to an overall understanding of the structure and mechanism of RNA polymerases. Nevertheless, crystal structures provide only static snapshots without information about the dynamics and it remains unknown if the proposed mechanisms are energetically feasible. Computational studies can fill the gaps between the structural snapshots and evaluate the energetics. Recent theoretical studies have been summarized[8,9] with special focus on the challenges multi subunit RNA polymerases pose to computational approaches. The nucleotide addition cycle[10,11] starts with a nucleotide triphosphate (NTP) entering the active site. If it matches the DNA template, the active site closes and the NTP is added to the growing RNA chain. After nucleotide addition, the resulting pyrophosphate is released, the active site opened and the polymerase is translocated by one nucleotide along the DNA. During nucleotide addition, the RNA polymerase active center efficiently discriminates against non-complementary NTPs and against complementary 2'- and 3'-deoxy NTPs (dNTPs).[12] Non-complementary NTPs are discriminated already in the open active site, because non Watson-Crick base pairs do not lead to closure of the active center. Although 3'-dNTPs can be discriminated, 3'-dNTPs do usually not occur and addition would lead to termination of the RNA strand and not to a usable result. 2'-dNTPs are discriminated after the active center closes due to an incorrect position that slows down the nucleotide addition reaction allowing expulsion of the incorrect 2'-dNTP.[12,13] Our study focuses on this discrimination in the active site (**Paper 1**). We identify the determinant factor, which is responsible for the incorrect positioning of 2'-dNTPs, and quantify its influence on the reaction profile, complementing recent experimental work.[12] We employed a QM/MM scheme that combines an accurate quantum mechanical (QM) description of the active site with a fast molecular mechanical (MM) description of the remaining system. Converging the height of the reaction barrier with the size of the QM region, we found 800 QM atoms to be necessary for an accurate description. Other studies[14–21] also showed the importance of converged QM sizes in the QM/MM approach, employing different QM/MM schemes.

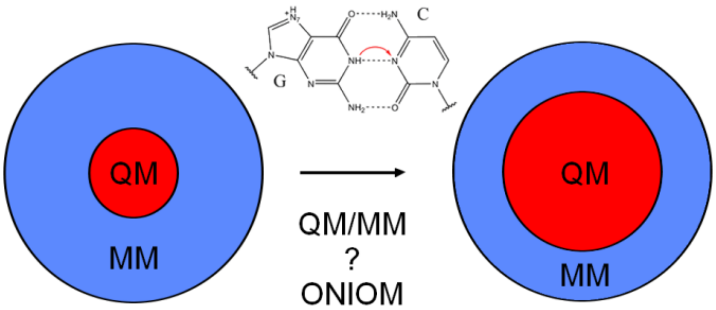


Figure 2. Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA.

In the second part of this thesis, the influence of different coupling (subtractive or additive) and embedding schemes (mechanical or electrostatic) on the QM size convergence behavior is investigated (**Paper 2**). We show that the choice of the embedding scheme is critical and also consider an improved variant of mechanical embedding[22] with QM enhanced charges. As test system, the proton transfer within a protonated DNA base pair has been chosen. It has been shown repeatedly, that an increased QM region has significant influence: For similar systems it was suggested to include the surrounding base pairs,[23–25] counter ions,[26] or the first hydration layer.[27–29] We systematically studied the QM size convergence behavior and now recommend that further studies of such proton transfers in DNA should include at least 5 base pairs and 5 Å of solvent into the QM region for reducing the error due to the QM size below 1 kcal/mol.

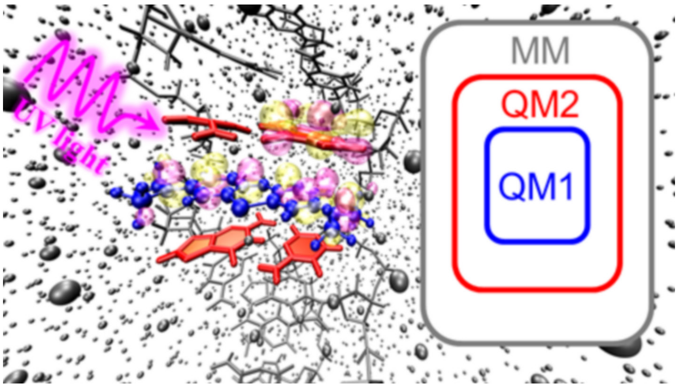


Figure 3. Effect of DNA Environment on Electronically Excited States of Methylene Blue in a QM/QM/MM Scheme.

In the last part of this thesis, the influence of an additional QM layer on vertical excitation energies is investigated for the example of methylene blue intercalated in DNA (**Paper 3**). Methylene blue is a phenothiazinium dye that has shown promising results in photodynamic therapy for tumor[30–32] and microbial infection treatment.[33,34] Photodynamic therapy is a minimal invasive and minimal toxic treatment strategy employing a photosensitizer that can be locally activated by exposure to light. The activated photosensitizer excites molecular oxygen, which damages the surrounding cell, eventually leading to cell death. Recently, it has been shown that the environment has a significant influence on the intersystem crossing mechanism of methylene blue. Moreover, intersystem crossing is enhanced in methylene blue, intercalated into DNA.[35] Accurate calculation of excitation energies usually requires multi-configurational methods (like CASSCF, CASPT2) that are accurate but only applicable to limited QM sizes.[36] In a QM/MM scheme, the whole surrounding environment is therefore treated only on the MM level. One advantage of the subtractive ONIOM scheme is the possibility to add additional layers with different levels of theory to the description of the system. This can be advantageous if the necessary level of theory does not permit the use of converged QM sizes. In such a case, an additional layer can be introduced between the low-level MM and the high-level QM regions with a medium-level theory that is less accurate but significantly faster than the high-level method. In contrast to the MM description, an additional QM layer includes polarization effects and charge transfer, facilitating more accurate models. Even chemical reactions can be described. For the case of methylene blue, the influence of an additional TD-DFT layer on the vertical excitation energies is investigated, considering two different binding modes. The energy shifts are separated into components due to geometry optimization, polarization, and electronic coupling, the latter being the dominant part in the case of intercalation.

Theory

This section is meant to give a short introduction and overview of the methods employed in the different parts of this thesis. A more detailed description can be found in numerous reviews.[36–42]

Force Fields. A force field (FF) used in molecular mechanics (MM) consists of a set of parameters to calculate the potential energy of a system of atoms. The parameters are usually derived from quantum mechanical calculations or fitted to experimental data. As in the current AMBER (Assisted Model Building with Energy Refinement) force field,[43] the generic form of the potential energy functional V consists of terms describing bonds, angles, dihedrals, van der Waals interactions, and Coulomb interactions (equation 1). It can be extended by special terms e.g. for 1-3 / 1-4 / 1-5 interactions, H-bonds, protein backbone / side chains to improve standard parameters.

$$
\begin{aligned}
V_{\text{Amber}} = {} & \sum_{\text{bonds}} k_b\left(r - r_{eq}\right)^2 + \sum_{\text{angles}} k_a\left(\theta - \theta_{eq}\right)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] \\
& + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{i<j}\left[\frac{q_i q_j}{\epsilon_r R_{ij}}\right]
\end{aligned}
\tag{1}
$$

k_b = bond force constant
r = bond length
r_{eq} = equilibrium bond length
k_a = angle force constant
θ = angle
θ_{eq} = equilibrium angle
V_n = torsion barrier
n = periodicity
ϕ = torsion angle
γ = phase
$A_{ij} = \epsilon_{ij} R_{min}^{12}$
$B_{ij} = 2\epsilon_{ij} R_{min}^{6}$
ϵ_{ij} = depth of the potential well
$R_{min} = R_i + R_j$
R_i = van der Waals radius of atom i
R_{ij} = distance separating the two atoms i and j
q_i = partial charge of atom i
ϵ_r = relative permittivity

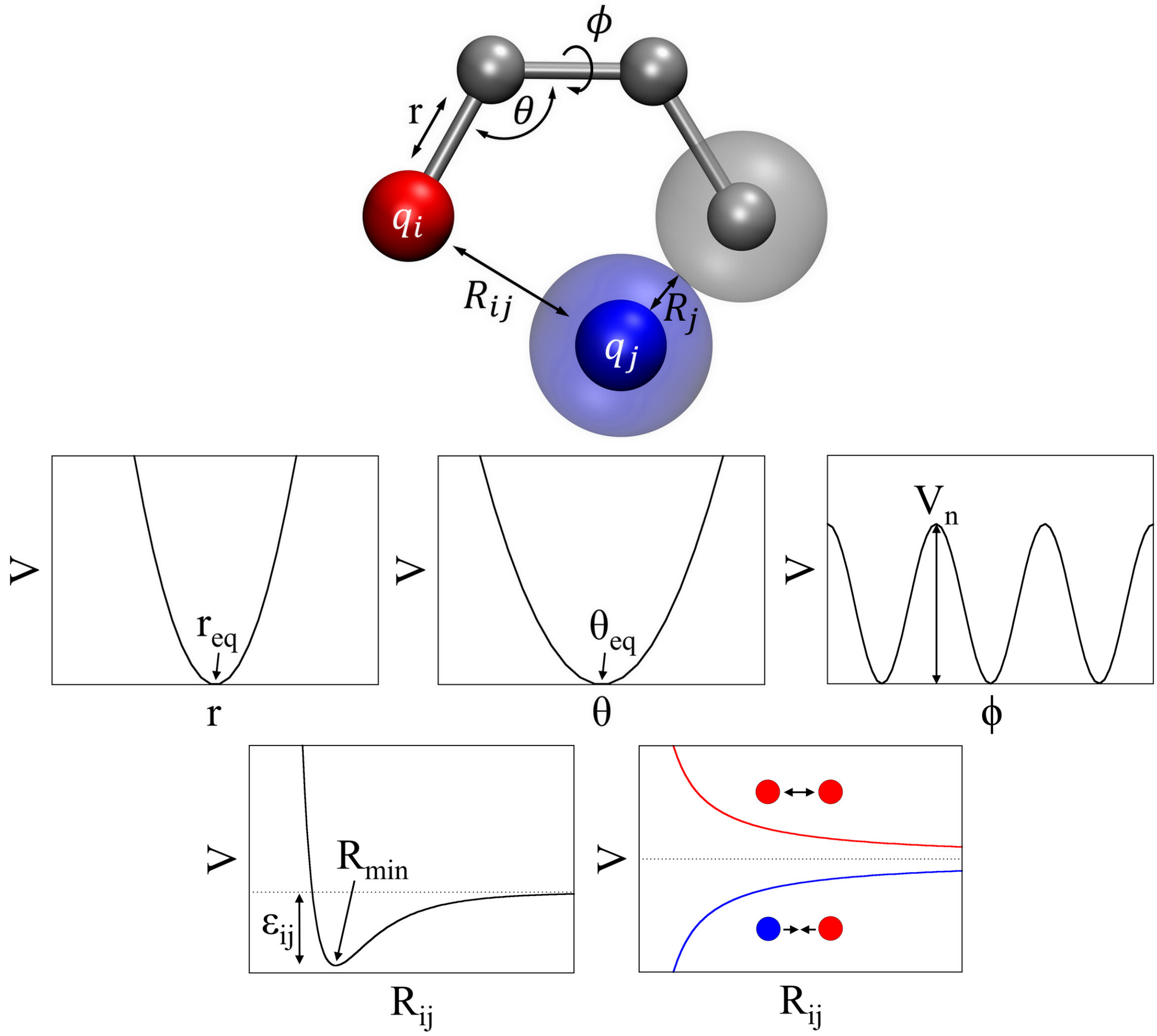


Figure 4. Visualization of the five terms of equation 1. The potentials of the bonded terms (bond, angle, torsion) are shown in the middle and the potentials of the non-bonded terms (van der Waals, electrostatic) are shown on the bottom.

The first three terms of equation 1 represent bonded terms that are visualized in the middle of Figure 4. Bonds and angles are described by harmonic terms that are parametrized by a force constant and the equilibrium value. Torsions are parametrized by the torsion barrier V_n between two minima, the periodicity n, e.g., three in the case of a methyl group, and a phase angle that is usually chosen so that terms with positive V_n have minima at 180°. To ensure planarity of four atoms that are not linearly bonded, like a carbonyl group or aromatic rings, "improper" torsion terms are applied. These describe the out-of-plane movement of the central atom. The remaining two terms describe non-bonded interactions that are visualized at the bottom of Figure 4. Van der Waals interactions are modelled by a standard 6-12 Lennard-Jones potential. The parameters A_{ij} and B_{ij} include the depth of the potential well ε_{ij} and the van der Waals radii R_i and R_j of the atoms i and j. The interaction of charged atoms is described by Coulombs law and depends on the partial charges q_i and q_j.

The main advantage of force fields is their speed, which allows modeling and all-atom molecular dynamics (MD) simulations even of large bio- and macromolecules.[44–46] MD simulations usually start from a high-resolution crystal structure. Any important parts that are not resolved need to be modeled and mutations / agents that allowed crystallization need to be reverted / removed. Among the disadvantages of classical force fields are the fixed topology that prevents the description of chemical reactions, fixed charges and the inability to describe excited states. However, reactive[47,48] and polarizable[49,50] force fields are developed to increase the applicability of force fields.

Quantum Mechanical methods. In contrast to parametrized force fields, *ab initio* quantum mechanical (QM) methods are parameter free and depend only on natural constants. The goal is the solution of the Schrödinger equation employing different assumptions and approximations to reduce the computational effort, which leads to a hierarchy of methods with increasing accuracy and cost. The Hartree-Fock (HF) method is one of the basic methods with relatively low computational effort and many approximations: the Born-Oppenheimer approximation allows to ignore the movement of nuclei for the calculation of the electronic wave function, relativistic effects and electron correlation for electrons of opposite spin are neglected and the electronic wave function is assumed to be describable by a single Slater determinant. The HF equations are solved iteratively until the electronic wave function is optimized with respect to the energy and a self-consistent field (SCF) is reached.

Another popular possibility to calculate the electronic structure of a system is based on the electron density instead of a wave function. The basis of density functional theory (DFT) are the Hohenberg-Kohn theorems[51] that state that the exact energy is a functional of the electron density that is subject to the variational principle. Therefore the number of variables reduces to the 3 space coordinates of the electron density instead of the 3N space coordinates of all electrons in the system. Unfortunately, the exact form of the density functional is not known and requires approximations. A large variety of approximate density functionals has been designed for various purposes and the most suitable DFT functional for a given problem can only be benchmarked by careful comparison to experimental data or higher level QM methods.

All of the mentioned methods are well suited to describe the ground state of a given system but are unable to describe excited states. Such calculations need to explicitly include the promotion of electrons into virtual orbitals. In principle, this can be done for example by including all possible single excitations into the HF wave function. Minimization of the energy leads to the Configuration Interaction Singles (CIS) approach. Due to Brillouin's theorem, the ground state and singly excited determinants do not interact, which means that CIS still does not account for electron correlation. Nevertheless, CIS excitation energies are reasonable when the ground state and the excited state are well described by a single configuration. However, this is often not the case as electron correlation plays an important role.[36] Although there is no clear separation, electron correlation is often divided into static and dynamic correlation. *Static correlation* results from, typically only a few, additional (near-) degenerate configurations that contribute strongly to the nature of the wave function. It can be described, e.g., by multi-configurational SCF methods like the complete active space (CAS) SCF approach,[52,53] which includes all symmetry allowed configurations from a chosen "active" orbital subspace. *Dynamic correlation* is directly related to dynamics because it originates from the instantaneous interaction of electrons upon movement, which is not described by HF theory. It typically involves a large number of configurations, each with a small weight, and can be approximated by, e.g., perturbation theory. Among the multi-reference perturbation methods, CASPT2[54,55] has become very popular. It is based on Møller-Plesset perturbation theory applied on a CASSCF reference wave function and includes all double excitations.

Despite the rapid development in computer speed and computational power, the size of typical biomolecular systems is still too large to be described at any level of *ab initio* theory. To overcome the size limitations of QM calculations, the idea of the QM/MM method[38,56] is to divide the system into a QM region, containing the most important part of the system including the chemical reaction center, and an MM region, containing the rest of the system. In this way, the accuracy of QM and its ability to describe changes in the topology are combined with the speed of MM and its ability to describe a large number of conformations. The partitioning of the system depends mainly on chemical intuition and experience; however, the result can be verified by systematically increasing the QM region.

In many systems, the QM/MM boundary cuts through a covalent bond, resulting in a dangling bond in the QM calculation. A simple way of solving this problem is to saturate this bond with a link atom, which is usually a hydrogen atom for the sake of simplicity, but can also be a halogen atom to model a polarized bond. In general, the QM/MM border must only cut through single bonds, preferably non-polarized single bonds, which is often impossible in practice. Cutting through conjugated systems should be avoided, because the link atom would be an insufficient approximation. Depending on the coupling scheme, force field parameters of the link atom are necessary.

Coupling schemes. QM and MM regions can be coupled in different ways, although the most common coupling schemes are either additive or subtractive (Figure 5). Subtractive schemes require an MM calculation of the full system (real system), a QM calculation of the inner part (model system), which is added, and also an MM calculation of the inner part, which is subtracted to avoid double counting. It can also be thought of as a QM calculation of the inner part with the effect of the remaining system extrapolated at the MM level. The most popular subtractive scheme is the "Our own N-layered Integrated molecular Orbital and molecular Mechanics" (ONIOM) method[57] by Morokuma and coworkers.

$$E^{\mathrm{ONIOM}} = E_{\mathrm{real}}^{\mathrm{MM}} + E_{\mathrm{model}}^{\mathrm{QM}} - E_{\mathrm{model}}^{\mathrm{MM}}$$

This generalization allows to combine any methods, including QM/QM, and can be extended to more than two layers. By design, the interactions between the layers are described on the lower level without any additional terms. As the model system is also calculated at the MM level, parameters for the link atoms are necessary, which allow to explicitly correct possible artifacts due to the introduction of link atoms at the QM level. However, this requires accurate parameters of these link atoms.

Additive coupling schemes include a QM calculation of the inner part, an MM calculation of the outer part and the evaluation of an explicit coupling term.

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}$$

$E_{\mathrm{QM/MM}}$ includes van der Waals interactions and classical bonding terms in the case of covalent bonds between QM and MM, although some of these force field terms are removed because they are already included in the QM part. In contrast to a subtractive scheme, both regions are calculated only once, force field parameters of QM atoms are only required at the QM/MM border[38] and link atoms are never part of the MM system.

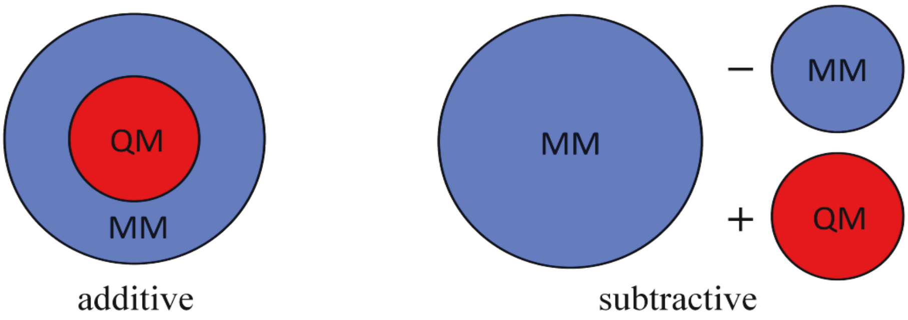


Figure 5. Schematic partitions by different coupling schemes.

Embedding schemes. In a QM/MM scheme, QM and MM region can interact in different ways, depending on the embedding scheme. The simplest case is mechanical embedding, which treats the electrostatic interactions on the lower level (MM) and neglects mutual polarization of the regions. Electrostatic embedding treats the electrostatic interactions at the higher level (QM) by including the partial charge of each MM atom as point charge into the one-electron Hamiltonian, polarizing the QM region (Figure 6).

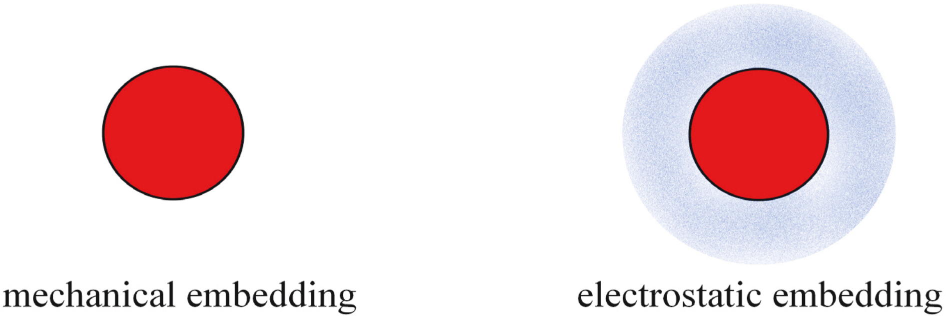


Figure 6. Different embedding schemes.

However, if the QM/MM border cuts through a covalent bond (Figure 7, top) and a link atom is introduced to saturate the QM region, the point charge of the next MM atom is very close (Figure 7, middle), leading to significant over-polarization. One way to handle the over-polarization of the QM region due to link atoms is to delete or scale the first layer(s) of corresponding point charges.[58–60] A more evolved scheme is to shift the charge of the first MM atom, which is connected to the QM atom, onto all connected MM atoms.[61,62] The charge shift leads to an artificial dipole moment, which can be corrected by the introduction of additional artificial dipoles at the position of the shifted charges (Figure 7, bottom).

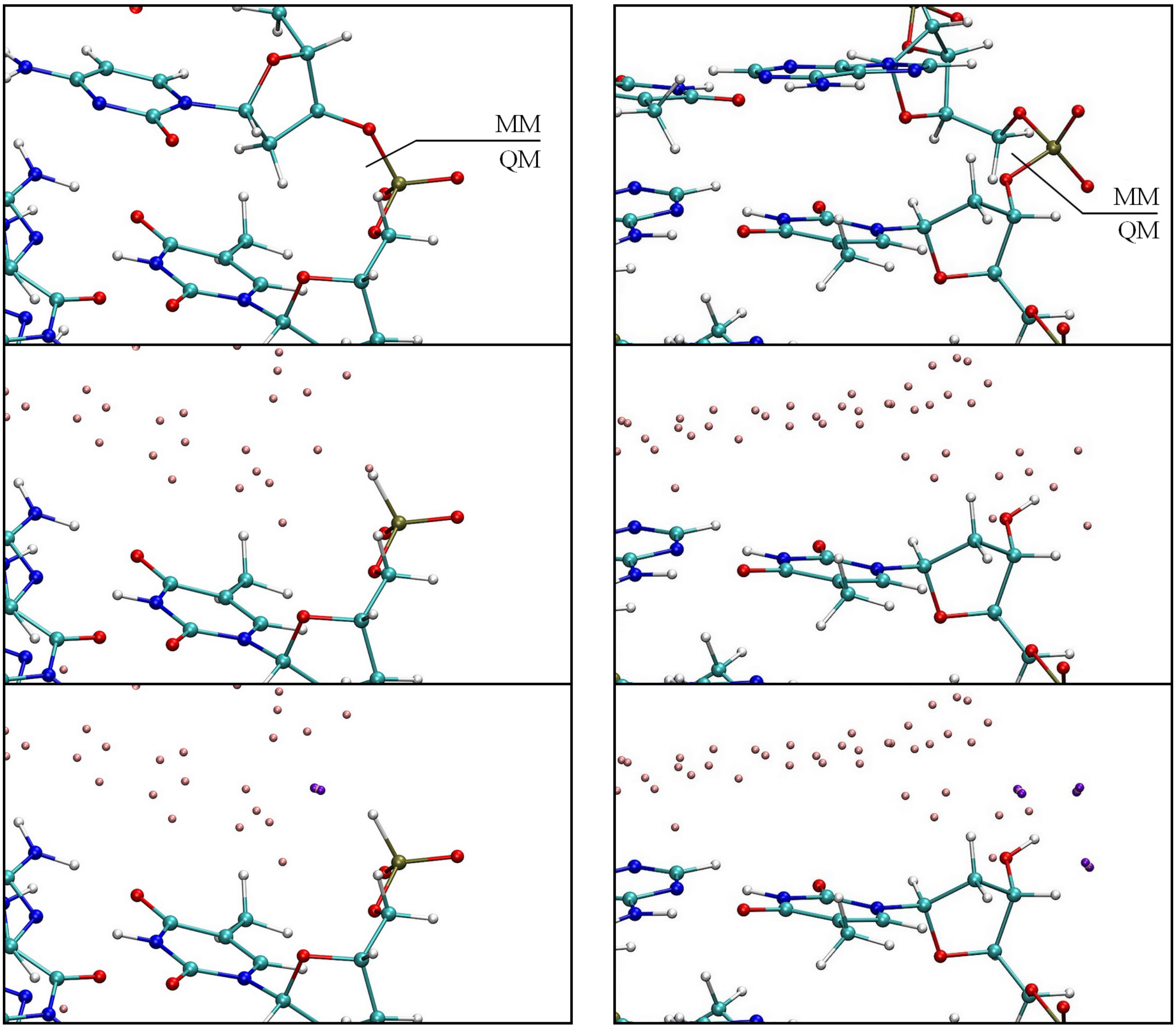


Figure 7. QM/MM border treatment with the charge shift scheme at the examples of 5'-end (left) and 3'-end (right) of a DNA strand. At the top, the QM/MM partitioning is shown. The picture in the middle shows the system from the perspective of the QM part, with the MM part consisting only of polarizing point charges (beige). The QM part is saturated with a link atom (hydrogen). At the bottom, the point charge of the MM atom that was very close to the link atom is shifted along the former MM bonds onto the next point charges. The resulting dipole moment is compensated by artificial dipoles at the same position (magenta).

The most advanced scheme would be polarizable embedding, which accounts also for polarization of the MM region and solves the mutual polarization iteratively until self-convergence. Until now, polarizable embedding is not employed regularly, because the advantage over electrostatic embedding is rather small compared to the enormous additional effort.[63] Also, e.g., overpolarization effects may be problematic.[64–66] Furthermore, they are currently not supported by most software packages.

3-layer ONIOM-EE. In principle, the ONIOM scheme allows combining any number of layers with any level of theory. Nevertheless, in the usual QM/MM partitioning two layers are enough: Most of the system is described at the MM (force field) level, which is appropriate for the description of geometries and classical dynamics, and only a small part of the system is described at the QM (*ab initio*) level to properly describe a chemical reaction. A third layer should only be introduced, if another level of theory is necessary to describe a certain part of the system that cannot be described otherwise. Excitation energies for example, often require time-dependent or multi-configurational methods that are limited to small QM regions.[36] Therefore a QM/QM/MM scheme with three layers is justified (Figure 8). The (local) excitation is treated on the highest level of theory in the smallest system. If the second layer is treated with a lower level *ab initio* method, polarization, charge transfer, and reactions can be described. Again, the remaining system is treated by a classical force field in the third layer. To include electrostatic embedding, Vreven et al.[67] suggested two possibilities:

$$E^{\text{ONIOM3-EE}} = E^{\text{v,QM(high)}}_{\text{model}} + E^{\text{v,QM(low)}}_{\text{int-model}} - E^{\text{v,QM(low)}}_{\text{model}} + E^{\text{MM}}_{\text{real}} - E^{\text{v,MM}}_{\text{int-model}} \tag{2}$$

$$E^{\text{ONIOM3-EEx}} = E^{\text{QM(high)}}_{\text{model}} + E^{\text{v,QM(low)}}_{\text{int-model}} - E^{\text{QM(low)}}_{\text{model}} + E^{\text{MM}}_{\text{real}} - E^{\text{v,MM}}_{\text{int-model}} \tag{3}$$

The superscript v indicates polarizing point charges taken from $E^{\text{MM}}_{\text{real}}$.

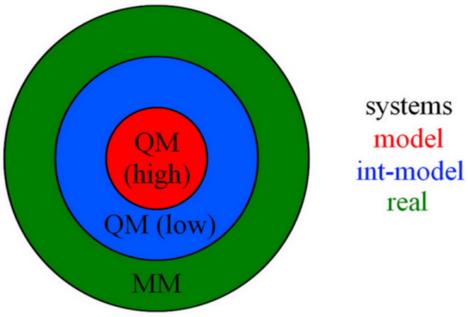


Figure 8. Partitioning in the 3-layer ONIOM scheme.

In the ONIOM3-EE scheme (equation 2), the model system and the int-model system are embedded into the point charges of the MM layer. In the ONIOM3-EEx scheme (equation 3), the point charges are excluded from the calculations of the model system, which is supposed to mimic the screening effect of the intermediate system on the model system. The authors found that it is necessary to include the point charges also in the calculations of the model system (ONIOM3-EE). When calculating excitation energies, Hall et al.[68] found it to be beneficial to include also the MM charges of the intermediate model into the calculations of the model system. However, static MM charges do not depend on the geometry of a system and cannot respond to structural changes or electronic excitations. Wanko et al.[69] employed a QM/QM/MM scheme with mutual polarization of the QM regions by iteratively fitted RESP charges. The effect on the excitation energy was only moderate, which was explained by the limited size of the outer QM region. Very recently, Biancardi et al.[70] investigated the influence of QM-derived charges on excitation energies and oscillator strengths in a two layer QM/QM-EE scheme. The main focus lay on an iterative ESP fitting scheme to account for link atoms. All tested variants performed similarly well, depending on the system. It was concluded that charges derived from the ground state are not enough and that the polarization response of the embedding point charges to electronic excitation should improve the results.

In our 3-layer ONIOM-EE scheme, the QM(high) layer is embedded in the MM point charges of the outer layer and the QM-derived charges of the intermediate layer, which depend on the electronic state (Figure 9). The over-polarization problem due to link atoms at the QM/QM and QM/MM borders is treated by the charge-shift scheme (Figure 7). In this 3-layer ONIOM-EE scheme, $\mathrm{QM}^{\mathrm{low}}_{\text{int-model}}$, $\mathrm{QM}^{\mathrm{high}}_{\mathrm{model}}$, and $\mathrm{QM}^{\mathrm{low}}_{\mathrm{model}}$ are usual QM calculations with polarizing point charges. $\mathrm{MM}_{\text{int-model}}$, however, is the non-standard case of an MM calculation with point charges that interact with the MM atoms but not with each other. Although the point charges are not polarizing the MM atoms, their inclusion is needed to avoid double counting, because the electrostatic QM-MM interaction is part of $\mathrm{MM}_{\mathrm{real}}$ and $\mathrm{QM}^{\mathrm{low}}_{\text{int-model}}$. $\mathrm{MM}_{\text{int-model}}$ also needs to include the link atoms, shifted charges, and additional dipoles, introduced by the charge shift scheme (Figure 7). All of these are non-standard and require additional parameters.

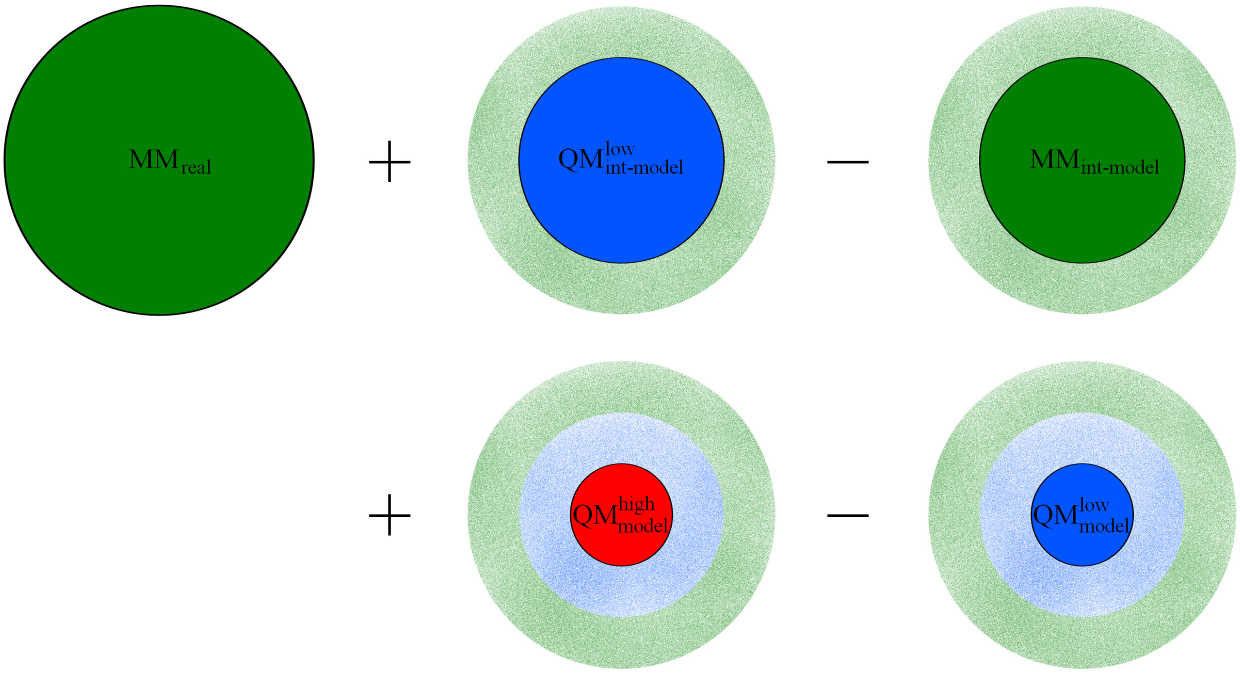


Figure 9. Partitioning in the 3-layer ONIOM-EE (QM/QM/MM) setup. Dots around the black circle represent point charges. Green corresponds to MM, and blue to QM(low).

For geometry optimizations, the required gradients can also be calculated with the ONIOM formalism, which is in the 3-layer-EE case:[67,71]

$$\begin{aligned}\frac{E^{\mathrm{ONIOM}}}{\partial\lambda} = \frac{E^{\mathrm{MM}}_{\mathrm{real}}}{\partial\lambda} &+ \frac{E^{\mathrm{v,QM(low)}}_{\text{int-model}}}{\partial\lambda^{\mathrm{QM(low)}}_{\text{int-model}}}\boldsymbol{J}^{\mathrm{QM(low)}}_{\text{int-model}} - \frac{E^{\mathrm{v,MM}}_{\text{int-model}}}{\partial\lambda^{\mathrm{MM}}_{\text{int-model}}}\boldsymbol{J}^{\mathrm{MM}}_{\text{int-model}} \\ &+ \frac{E^{\mathrm{v,QM(high)}}_{\mathrm{model}}}{\partial\lambda^{\mathrm{QM(high)}}_{\mathrm{model}}}\boldsymbol{J}^{\mathrm{QM(high)}}_{\mathrm{model}} - \frac{E^{\mathrm{v,QM(low)}}_{\mathrm{model}}}{\partial\lambda^{\mathrm{QM(low)}}_{\mathrm{model}}}\boldsymbol{J}^{\mathrm{QM(low)}}_{\mathrm{model}}\end{aligned} \tag{4}$$

The Jacobian $\boldsymbol{J}$ is used to convert the coordinate systems to that of the real system. If link atoms are present, $\boldsymbol{J}$ must be used to project the gradients of the link atoms onto the respective QM and MM atoms.[71]

Instead of the ONIOM-EE scheme outlined above, Ryde and coworkers[19,22,72–74] employed a different scheme that avoids double counting of the electrostatic interactions between QM and MM region (Figure 10). Instead of including the electrostatic interactions in the MM calculation of the QM region and subtracting it, the charge of all atoms in the QM region is set to zero in both MM calculations. Therefore, no special point charges, shifted charges, or dipoles are necessary in the MM calculation of the QM region and only the link atoms require additional parameters. The drawback of this scheme is that link atoms are now treated differently in the QM and MM calculation, leading to artifacts. In the QM calculation, link atoms have a charge that interacts with the QM region and the surrounding point charges, whereas this is not the case in the MM calculation. Furthermore, gradients of atoms in the MM region will be incorrect, because they are calculated only in the MM calculation of the full system (see equation 4), which lacks the interaction with the charges of the atoms in the QM region.

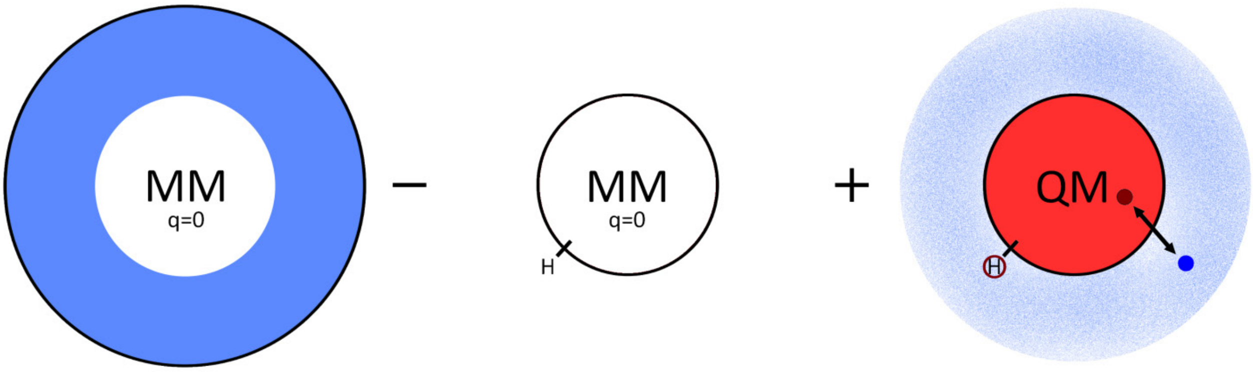


Figure 10. Alternative subtractive QM/MM scheme with modified electrostatics in the MM calculations to avoid double counting. The electrostatic interactions between the QM and MM region is only part of the QM calculation.

Reaction paths. Although the calculation of reaction barriers is a very important application of quantum mechanics, it is not straightforward and there are different approaches and levels of approximation. Probably the simplest and most intuitive approach is the "adiabatic mapping" or "drag" method. In this method, a reaction coordinate is chosen and the system is dragged along this reaction coordinate by a series of restrained geometry optimizations. During the geometry optimization, all other degrees of freedom are allowed to relax adiabatically, mapping an energy to each restraint along the reaction coordinate. This scheme can in principle provide the correct result but often fails because the chosen reaction coordinate is not good enough to describe the reaction properly.[41,75,76] In such cases, discontinuities in the resulting energy profile are frequently observed due to sudden structural relaxation of the environment, which was not included in the chosen reaction coordinate. The relaxation might even include atoms far away from the active site, leading to changes in the relative energy that are unrelated to the investigated reaction. This can be avoided by freezing the coordinates of atoms that are "far" away from the active site. Another indication of an erroneous reaction coordinate is a hysteresis of the energy profile when calculating the backward reaction, which is often caused by a reorganization of the environment.

Other schemes have been developed to find real minimal energy paths (MEPs) between two structures. Conjugate peak refinement,[77] for example, finds a minimal energy path with true saddle points without applying constraints to drive the reaction. An initial guess is extended and refined until a MEP is reached. Chain-of-states methods employ a number of images (also called states) to connect the start and end structure. The images are held together by constraints or restraint, depending on the method. In the case of the "nudged elastic band" (NEB) method,[75] adjacent images are connected by harmonic spring forces that ensure an equal distribution along the reaction coordinate. The whole chain is then minimized simultaneously until it converges on a MEP.

Free energy differences. When compared to energy barriers obtained from kinetic studies, MEPs completely neglect changes in the entropy, that is, the dynamics of the system. In classical force field dynamics, the method of choice to calculate entropic contributions is umbrella sampling (Figure 11a). The coordinate ξ, connecting two geometries, is divided into overlapping windows, which are then sampled individually with a harmonic biasing potential applied. If the histograms of these individual MD simulations overlap sufficiently, they can be analyzed, combined and the biasing potential removed by methods like the "Weighted Histogram Analysis Method"[78] or "Umbrella Integration".[79] The result is an unbiased probability distribution and the free energy change along the chosen coordinate can be calculated. The thermodynamic integration method (Figure 11b) also divides the coordinate ξ into windows but calculates the mean force on this coordinate. The free energy change is obtained by integration over the canonical average of the force constraints. Thermodynamic integration is equivalent to umbrella sampling in the limit of a strong biasing potential (constraint), analyzed with Umbrella Integration.[79] When calculating the free energy change along a reaction coordinate, quantum mechanical methods are needed to describe bond formation and breakage. Thermodynamic integration and umbrella sampling rely on an extensive sampling of the phase space, which is currently not feasible with QM methods. Even MD simulations employing only small QM regions within the QM/MM approach are often impractical. The QM/MM free energy perturbation (FEP) method[42] aims to avoid the expensive sampling of the QM by separating QM and MM contributions to the free energy difference. The entropy change within the QM part is approximated based on numeric frequency calculations of educt, transition state and product geometries, whereas the MM contribution is calculated with the FEP approach. The QM/MM-FEP approach (Figure 11c) starts from a series of geometries along the reaction coordinate, obtained, e.g., by adiabatic mapping or NEB. For each geometry, all QM atoms are replaced by atomic charges, fitted from the electrostatic potential of the QM calculation. The MM part of each geometry can be perturbed by the ESP charges of the next and/or the previous geometry along the reaction coordinate λ, leading to an energy difference due to forward and/or backward perturbation. Now, all perturbed systems are sampled by classical MD simulations, with the ESP charges fixed in position, and the free energy difference of the MM part can be calculated between adjacent structures. Summation of the free energy differences along the reaction coordinate results in the total free energy difference.

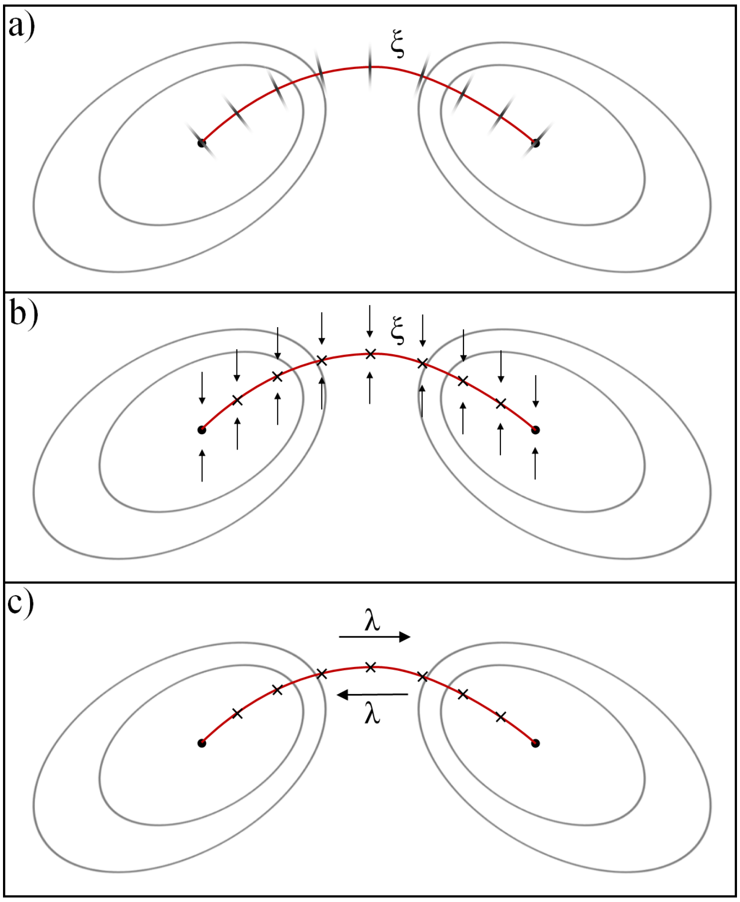


Figure 11. Sampling of the reaction coordinate (red line) between two states on an arbitrary potential energy surface by different methods. a) Umbrella Sampling: In each window, the system, restraint by a harmonic potential, is mainly sampled perpendicular to the reaction coordinate. b) Thermodynamic Integration: The reaction coordinate is constrained to different values and the mean force is sampled at each point. c) Free Energy Perturbation: The system is sampled along the reaction coordinate, perturbed by the succeeding (forward perturbation) or preceding (backward perturbation) geometry.

QM size convergence. Any result obtained from small QM regions, which have been obtained by chemical intuition, should be confirmed by a QM region with converged size. When the QM size is increased, the description of residues changes from MM to QM, allowing polarization and electronic effects. Residues that are close to the active site usually have the largest effect and are the first choice when choosing the QM region. Nevertheless, it has been shown that also other residues can be important and an increased QM region might even change the result qualitatively.[14,20,22,80,81] The most straightforward way is to systematically increase the QM size by whole residues based on the distance of the closest atom to the active site, or to a special selection of atoms, until there is no further change in the result. Other ways to estimate the importance of single residues have been suggested but have not received much attention, yet.[80,82,83] Currently there is no universal scheme to estimate the necessary QM size reliably. When the converged QM size has been found, re-optimization of the structures employing the converged QM region is necessary to determine its influence on the result. In the case of reaction profiles, this means that the whole reaction path should be re-calculated. In practice, it is usually enough to re-optimize educt-, product- and all intermediate minima-/maxima-structures along the reaction coordinate. Unfortunately, the optimization of large QM regions can result in structural changes far away from the active site, e.g., solvent rearrangement, leading to different local minima of structures along the reaction coordinate. These different local minima often have a large or even dominating impact on the reaction profile. To prevent such unrelated changes of the environment due to the increased QM region, geometry optimizations with a large, converged QM size can be limited to the active site.

Conclusions

In this work, the influence of complementary 2'-dNTPs on the nucleotide addition reaction of RNA Polymerase II, employing a QM/MM approach was investigated. To study the difference in the reaction of RNA Pol II with ATP and with dATP, we chose the crystal structure that shows the most interacting residues with the OH2' group of the NTP and found only one of the suggested mechanisms to be feasible for this crystal structure. We identified Arg446 as determinant factor, which is responsible for the incorrect positioning of 2'-dNTPs, and quantified its influence on the reaction profile, complementing recent experimental work. In the case of 2'dATP, the slightly different interaction with Arg446 results in a conformational change that leads to an increase of the reaction barrier by 13 kcal/mol, finally explaining how a single hydrogen bond can account for the high selectivity. These findings reveal the crucial role of Arg446 in the recognition of the 2'-OH group that could not be assessed by mutation studies. To verify our results, we estimated the free energy change along the reaction coordinate by a QM/MM-FEP approach and converged the height of the reaction barriers with the size of the QM region. For an accurate description at least 800 atoms need to be included into the QM region.

Since QM/MM calculations can be performed with different coupling and embedding schemes, we systematically investigated their influence on the QM size convergence at the example of a proton transfer within DNA. While the choice of the coupling scheme between the QM and the MM region has no significant influence on the size of the converged QM region, we found the choice of the embedding scheme to be crucial. No embedding (pure QM) can lead to seemingly fast convergence that becomes apparently wrong only with very large QM regions. Mechanical embedding, which is often used in the ONIOM approach, can result in slow convergence and misleading energies. In the investigated case, only electrostatic embedding leads to a fast and reliable convergence of the QM size. For further studies of proton transfer processes in DNA, we recommend to include two adjacent base pairs on each side and at least 5 Å of solvent around these base pairs into the QM region.

For the example of methylene blue interacting with DNA in two different binding modes, we investigated the influence of an additional QM layer in QM/MM calculations on vertical excitation energies. Although the QM/QM/MM scheme, we employed, did not improve the accuracy compared to the experiment, we could show the influence of the additional layer and why a better description did not improve the overall accuracy for the present system. We split the energy shifts into an electrostatic contribution of the polarizing environment and an electronic-coupling component due to the increased QM size. While polarization and geometry optimization had only small and partly opposite effects on the vertical excitation energies, we expect this influence to be more pronounced for other chromophores with more diffuse excitations. The intercalated methylene blue, however, showed a significant red-shift of the bright $S_1(\pi_1\pi_1^*)$ state due to electronic-coupling to a charge-transfer state of the adjacent nucleobase, that can only be taken into account by the increased QM layer.

This work demonstrates the influence of QM size convergence at the examples of a chemical reaction in a large enzyme-DNA complex, different computational approaches, and excited state calculations on a dye molecule bound to DNA. The need to converge the QM size employed in QM/MM calculations to achieve accurate results is emphasized. Nevertheless, big QM regions pose a special challenge even to modern quantum chemical software. However, GPU accelerated computing shows a promising way to reduce computation time and to further extend the scope of application.

Bibliography

(1) Cramer, P. Multisubunit RNA Polymerases. *Curr. Opin. Struct. Biol.* **2002**, *12*, 89–97.

(2) Nikiforov, V. G. The RNA Polymerase Structure – Function Study (1962 – 2001). *Mol. Biol.* **2002**, *36*, 143–152.

(3) Cramer, P.; Armache, K.-J.; Baumli, S.; Benkert, S.; Brueckner, F.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jasiak, A. J.; Jawhari, A.; Jennebach, S.; Kamenski, T.; Kettenberger, H.; Kuhn, C.-D.; Lehmann, E.; Leike, K.; Sydow, J. F.; Vannini, A. Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **2008**, *37*, 337–352.

(4) Steitz, T. A. A Mechanism for All Polymerases. *Nature* **1998**, *391*, 231–232.

(5) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48*, 220–228.

(6) Sainsbury, S.; Bernecky, C.; Cramer, P. Structural Basis of Transcription Initiation by RNA Polymerase II. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 129–143.

(7) Hantsche, M.; Cramer, P. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angew. Chemie Int. Ed.* **2016**, *55*, 15972–15981.

(8) Wang, B.; Feig, M.; Cukier, R. I.; Burton, Z. F. Computational Simulation Strategies for Analysis of Multisubunit RNA Polymerases. *Chem. Rev.* **2013**, *113*, 8546–8566.

(9) Pardo-Avila, F.; Da, L.-T.; Wang, Y.; Huang, X. Theoretical Investigations on Elucidating Fundamental Mechanisms of Catalysis and Dynamics Involved in Transcription By Rna Polymerase. *J. Theor. Comput. Chem.* **2013**, *12*, 1–16.

(10) Brueckner, F.; Ortiz, J.; Cramer, P. A Movie of the RNA Polymerase Nucleotide Addition Cycle. *Curr. Opin. Struct. Biol.* **2009**, *19*, 294–299.

(11) Cheung, A. C. M.; Cramer, P. A Movie of RNA Polymerase II Transcription. *Cell* **2012**, *149*, 1431–1437.

(12) Yuzenkova, Y.; Bochkareva, A.; Tadigotla, V. R.; Roghanian, M.; Zorov, S.; Severinov, K.; Zenkin, N. Stepwise Mechanism for Transcription Fidelity. *BMC Biol.* **2010**, *8*.

(13) Wang, B.; Opron, K.; Burton, Z. F.; Cukier, R. I.; Feig, M. Five Checkpoints Maintaining the Fidelity of Transcription by RNA Polymerases in Structural and Energetic Details. *Nucleic Acids Res.* **2015**, *43*, 1133–1146.

(14) Sumowski, C. V.; Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **2009**, *113*, 11734–11741.

(15) Sumowski, C. V.; Schmitt, B. B. T.; Schweizer, S.; Ochsenfeld, C. Quantum-Chemical and Combined Quantum-Chemical/Molecular-Mechanical Studies on the Stabilization of a Twin Arginine Pair in Adenovirus Ad11. *Angew. Chemie - Int. Ed.* **2010**, *49*, 9951–9955.

(16) van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010**, *114*, 11303–11314.

(17) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8*, 3793–3803.

(18) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8*, 2260–2271.

(19) Sumner, S.; Söderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4205–4214.

(20) Liao, R.-Z.; Thiel, W. Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comput. Chem.* **2013**, *34*, 2389–2397.

(21) Blank, I. D.; Sadeghian, K.; Ochsenfeld, C. A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination Within the Active Site. *Sci. Rep.* **2015**, *5*, 10369.

(22) Hu, L.; Söderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

(23) Matsui, T.; Sato, T.; Shigeta, Y.; Hirao, K. Sequence-Dependent Proton-Transfer Reaction in Stacked GC Pair II: The Origin of Stabilities of Proton-Transfer Products. *Chem. Phys. Lett.* **2009**, *478*, 238–242.

(24) Matsu, T.; Sato, T.; Shigeta, Y. Sqeuence Dependent Proton-Transfer Reaction in Stacked GC Pair I: The Possibility of Proton-Transfer Reactions. *Int. J. Quantum Chem.* **2009**, *109*, 2168–2177.

(25) Villani, G. Theoretical Investigation of the Coupling between Hydrogen Atoms Transfer and Stacking Interaction in Guanine-Cytosine Dimers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19242–19252.

(26) Chen, H. Y.; Kao, C. L.; Hsu, S. C. N. Proton Transfer in Guanine- Cytosine Radical Anion Embedded in B-Form DNA. *J. Am. Chem. Soc.* **2009**, *131*, 15930–15938.

(27) Zhanpeisov, N. U.; Leszczynski, J. Specific Solvation Effects on the Structures and Properties of Neutral and One-Electron Oxidized Formamidine-Formamide Complexes. A Theoretical Ab Initio Study. *J. Phys. Chem. A* **1999**, *103*, 8317–8327.

(28) Cerón-Carrasco, J. P.; Requena, A.; Zúñiga, J.; Michaux, C.; Perpète, E. A.; Jacquemin, D. Intermolecular Proton Transfer in Microhydrated Guanine-Cytosine Base Pairs: A New Mechanism for Spontaneous Mutation in DNA. *J. Phys. Chem. A* **2009**, *113*, 10549–10556.

(29) Cerón-Carrasco, J. P.; Zúñiga, J.; Requena, A.; Perpète, E. A.; Michaux, C.; Jacquemin, D. Combined Effect of Stacking and Solvation on the Spontaneous Mutation in DNA. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14584–14589.

(30) Mellish, K. J.; Cox, R. D.; Vernon, D. I.; Griffiths, J.; Brown, S. B. In Vitro Photodynamic Activity of a Series of Methylene Blue Analogues. *Photochem. Photobiol.* **2002**, *75*, 392–397.

(31) Orth, K.; Rück, A.; Stanescu, A.; Beger, H. G. Intraluminal Treatment of Inoperable Oesophageal Tumours by Intralesional Photodynamic Therapy with Methylene Blue. *Lancet* **1995**, *345*, 519–520.

(32) Orth, K.; Russ, D.; Beck, G.; Rück, A.; Beger, H. G. Photochemotherapy of Experimental Colonic Tumours with Intra-Tumorally Applied Methylene Blue. *Langenbeck's Arch. Surg.* **1998**, *383*, 276–281.

(33) Zolfaghari, P. S.; Packer, S.; Singer, M.; Nair, S. P.; Bennett, J.; Street, C.; Wilson, M. In Vivo Killing of Staphylococcus Aureus Using a Light-Activated Antimicrobial Agent. *BMC Microbiol* **2009**, *9*, 27.

(34) Ragàs, X.; Dai, T.; Tegos, G. P.; Agut, M.; Nonell, S.; Hamblin, M. R. Photodynamic Inactivation of Acinetobacter Baumannii Using Phenothiazinium Dyes: In Vitro and in Vivo Studies. *Lasers Surg. Med.* **2010**, *42*, 384–390.

(35) Nogueira, J. J.; Oppel, M.; González, L. Enhancing Intersystem Crossing in Phenotiazinium Dyes by Intercalation into DNA. *Angew. Chemie - Int. Ed.* **2015**, *54*, 4375–4378.

(36) González, L.; Escudero, D.; Serrano-Andrés, L. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem* **2012**, *13*, 28–51.

(37) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*.

(38) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 1198–1229.

(39) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals Se. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.

(40) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V; Li, X.; Ke, Z.; Liu, F.; Li, H.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(41) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A Practical Guide to Modelling Enzyme-Catalysed Reactions. *Chem. Soc. Rev.* **2012**, *41*, 3025–3038.

(42) Kästner, J.; Senn, H. H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction. *J. Chem. Theory Comput.* **2006**, *23*, 452–461.

(43) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. AMBER 2016. University of California, San Francisco 2016.

(44) Andoh, Y.; Yoshii, N.; Yamada, A.; Fujimoto, K.; Kojima, H.; Mizutani, K.; Nakagawa, A.; Nomoto, A.; Okazaki, S. All-Atom Molecular Dynamics Calculation Study of Entire Poliovirus Empty Capsids in Solution. *J. Chem. Phys.* **2014**, *141*, 165101.

(45) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74.

(46) Chavent, M.; Duncan, A. L.; Sansom, M. S. P. Molecular Dynamics Simulations of Membrane Proteins and Their Interactions: From Nanoscale to Mesoscale. *Curr. Opin. Struct. Biol.* **2016**, *40*, 8–16.

(47) Hartke, B.; Grimme, S. Reactive Force Fields Made Simple. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16715–16718.

(48) Han, Y.; Jiang, D.; Zhang, J.; Li, W.; Gan, Z.; Gu, J. Development, Applications and Challenges of ReaxFF Reactive Force Field in Molecular Simulations. *Front. Chem. Sci. Eng.* **2016**, *10*, 16–38.

(49) Baker, C. M. Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. *WIREs Comput. Mol. Sci.* **2015**, *5*, 241–254.

(50) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Polarizable Force Fields for Biomolecular Modeling. In *Reviews in Computational Chemistry Volume 28*; John Wiley & Sons, Inc, 2015; pp 51–86.

(51) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.

(52) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. M. A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach. *Chem. Phys.* **1980**, *48*, 157–173.

(53) Roos, B. O. The Complete Active Space SCF Method in a Fock-Matrix-Based Super-CI Formulation. *Int. J. Quantum Chem.* **1980**, *18*, 175–189.

(54) Andersson, K.; Malmqvist, P.-A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(55) Andersson, K.; Malmqvist, P.-A.; Roos, B. O. Second-Order Perturbation Theory with a Complete Active Space Self-Consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(56) Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chemie Int. Ed.* **2014**, *53*, 10020–10031.

(57) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(58) Singh, U. C.; Kollman, P. A. 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.* **1986**, *7*, 718–730.

(59) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. Combined Quantum Mechanical–molecular Mechanical Study of Catalysis by the Enzyme Phospholipase A2 : An Investigation of the Potential Energy Surface for Amide Hydrolysis. *J. Chem. Soc., Perkin Trans. 2* **1991**, No. 12, 2025–2032.

(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16. 2016.

(61) Sherwood, P.; De Vries, A. H. .; Collins, S. J.; Greatbanks, S. P.; Burton, N. a.; Vincent, M. a.; Hillier, I. H. Computer Simulation of Zeolite Structure and Reactivity Using Embedded Cluster Methods. *Faraday Discuss* **1997**, *106*, 79–92.

(62) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct. Theochem* **2003**, *632*, 1–28.

(63) Thellamurege, N. M.; Hirao, H. Effect of Protein Environment within Cytochrome P450cam Evaluated Using a Polarizable-Embedding QM/MM Method. *J. Phys. Chem. B* **2014**, *118*, 2084–2092.

(64) Chelli, R.; Schettino, V.; Procacci, P. Comparing Polarizable Force Fields to Ab Initio Calculations Reveals Nonclassical Effects in Condensed Phases. *J. Chem. Phys.* **2005**, *122*.

(65) Vosmeer, C. R.; Rustenburg, A. S.; Rice, J. E.; Horn, H. W.; Swope, W. C.; Geerke, D. P. QM/MM-Based Fitting of Atomic Polarizabilities for Use in Condensed-Phase Biomolecular Simulation. *J. Chem. Theory Comput.* **2012**, *8*, 3839–3853.

(66) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983–5013.

(67) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery Jr., J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory* **2006**, *2*, 815–826.

(68) Hall, K. F.; Vreven, T.; Frisch, M. J.; Bearpark, M. J. Three-Layer ONIOM Studies of the Dark State of Rhodopsin: The Protonation State of Glu181. *J. Mol. Biol.* **2008**, *383*, 106–121.

(69) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. Effect of Polarization on the Opsin Shift in Rhodopsins. 1. A Combined QM/QM/MM Model for Bacteriorhodopsin and Pharaonis Sensory Rhodopsin II. *J. Phys. Chem. B* **2008**, *112*, 11462–11467.

(70) Biancardi, A.; Barnes, J.; Caricato, M. Point Charge Embedding for ONIOM Excited States Calculations. *J. Chem. Phys.* **2016**, *145*, 224109.

(71) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct. Theochem* **1999**, *461–462*, 1–21.

(72) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. Proton Transfer at Metal Sites in Proteins Studied by Quantum Mechanical Free-Energy Perturbations. *J. Chem. Theory Comput.* **2008**, *4*, 985–1001.

(73) Hu, L.; Söderhjelm, P.; Ryde, U. Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations. *J. Chem. Theory Comput.* **2013**, *9*, 640–649.

(74) Li, J.; Farrokhnia, M.; Rulíšek, L.; Ryde, U. Catalytic Cycle of Multicopper Oxidases Studied by Combined Quantum- and Molecular-Mechanical Free-Energy Perturbation Methods. *J. Phys. Chem. B* **2015**, *119*, 8268–8284.

(75) Jonsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998; pp 385–404.

(76) Henkelman, G.; Jóhannesson, G.; Jónsson, H. Methods for Finding Saddle Points and Minimum Energy Paths. In *Theoretical Methods in Condensed Phase Chemistry*; Schwartz, S. D., Ed.; Springer Netherlands: Seattle, 2002; pp 269–302.

(77) Fischer, S.; Karplus, M. Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom. *Chem. Phys. Lett.* **1992**, *194*, 252–261.

(78) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(79) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: “umbrella Integration.” *J. Chem. Phys.* **2005**, *123*.

(80) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do Quantum Mechanical Energies Calculated for Small Models of Protein-Active Sites Converge? *J. Phys. Chem. A* **2009**, *113*, 11793–11800.

(81) Przybylski, J. L.; Wetmore, S. D. A QM/QM Investigation of the hUNG2 Reaction Surface: The Untold Tale of a Catalytic Residue. *Biochemistry* **2011**, *50*, 4218–4227.

(82) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O -Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381–11394.

(83) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM / MM Simulation. *J. Chem. Theory Comput.* **2017**.

Publications

Paper 1: “Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II”, S. Roßbach, C. Ochsenfeld, *J. Chem. Theory Comput.*, **2017**, *13*, 1699-1705.

JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

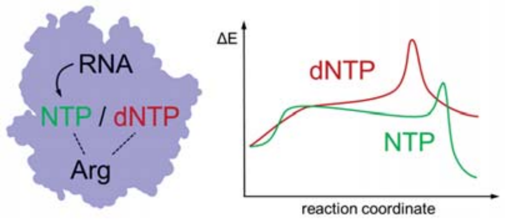
Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II

Sven Roßbach[†,‡] and Christian Ochsenfeld*[,†,‡]

[†]Chair of Theoretical Chemistry, Department of Chemistry and [‡]Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU Munich), Butenandtstrasse 7, D-81377 Munich, Germany

Supporting Information

ABSTRACT: Eukaryotic RNA polymerase II catalyzes the transcription of DNA into mRNA very efficiently and with an extremely low error rate with regard to matching base and sugar moiety. Despite its importance, little is known about how it discriminates against 2′-deoxy NTPs during the chemical reaction. To investigate the differences in the addition reactions of ATP and dATP, we used FF-MD and QM/MM calculations within a nudged elastic band approach, which allowed us to find the energetically accessible reaction coordinates. By converging the QM size, we found that 800 QM atoms are necessary to properly describe the active site. We show how the absence of a single hydrogen bond between the enzyme and the NTP 2′-OH group leads to an increase of the reaction barrier by 16 kcal/mol and therefore conclude that Arg446 is the key residue in the discrimination process.



INTRODUCTION

RNA polymerase II (Pol II) is the eukaryotic enzyme catalyzing the transcription of protein-coding genes into a complementary chain of ribonucleotides. When Pol II reads the DNA template and synthesizes the new RNA strand, a highly accurate transfer of genetic information is essential for the process of life. During transcription, Pol II discriminates against noncomplementary ribonucleotide triphosphates (nc-rNTPs) as well as against 2′- and 3′-deoxy nucleotide triphosphates (dNTPs) with a transcription error rate of 10^{-5} per nucleotide.[1] Possible discrimination in the preinsertion site has been discussed,[2] but the relevant residues have not been revealed.

In the insertion site, recognition of base and sugar is achieved by a sensitive network of hydrogen bonds between NTP and different parts of the active site. The trigger loop (TL), a mobile element that closes the active site after NTP entry, forms NTP contacts and was found to improve discrimination by kinetic selection[3−5] and/or catalysis.[6] Pol II discrimination against nc-rNTPs is mainly based on Watson−Crick base pairing. Mismatched NTPs are misaligned within the active site and collide with the closing TL.[7,8] It is not entirely clear how discrimination against 3′dNTPs is achieved. The mechanism might involve a contact to Gln1078, which is needed to reach a catalytically active conformation.[7] Discrimination against 3′dNTPs is much less effective than against 2′dNTPs, but incorporation would lead to strand termination and 3′dNTPs usually do not occur *in vivo*. 2′dNTPs on the other hand do occur regularly, as they form DNA, and discrimination against 2′dNTPs is very effective (∼50000).[7] Three residues have been identified in Pol II (and bacterial RNA Polymerase), that could be responsible for discrimination against 2′dNTPs via H-bond to 2′-oxygen: Asn479[3,9] (Asn β′737[2,7,10]), Arg446[3,9] (Arg β′704[11]), and Gln1078[9] (Gln β′1235[7]). Although there are mutation studies for the mentioned residues, the details of how a mutation affects discrimination remain unclear. The mutation might disrupt the sensitive H-bond network which leads to a slightly different orientation of other important residues, resulting in less effective discrimination, even though the mutated residue did not form a critical contact itself.[4,12] Therefore, mutation studies are typically not sufficient to explain the discrimination process on the molecular level or why a simple H-bond results in such selectivity. However, multiple steps of discrimination against 2′dNTPs have been discussed, and kinetic studies[3,7] as well as theoretical studies[13,14] revealed that selection against 2′dNTPs occurs mainly during the chemical step.

X-ray structures of reaction intermediates require modifications to stop the reaction at a specific step, possibly leading to artificial states. In order to provide reliable insights, the reaction was simulated and further investigated in computational studies. The mechanism has been assumed to follow the general two-ion mechanism also found for other nucleic acid polymerases:[15] The terminal $O^{3'}_{\mathrm{RNA}}$ is deprotonated by a general base to attack $P^{\alpha}_{\mathrm{NTP}}$, and the leaving pyrophosphate is protonated by a general acid. Previous theoretical studies aimed to elucidate the general base/acid and to validate the two-ion mechanism also for Pol II.

Zhu and Salahub[16] employed stirred ReaxFF-MD on a backbone model of 250 atoms solvated by 400 water molecules. They found similar barriers for $OH^{3'}$ deprotonation by a nearby

Received: February 14, 2017
Published: March 8, 2017

DOI: 10.1021/acs.jctc.7b00157
J. Chem. Theory Comput. 2017, 13, 1699−1705

water and by Asp485. The conclusion was a “double check mechanism”, deprotonating $O^{3'}$ of the terminal RNA by water during translocation. Asp485 should work as backup to deprotonate $O^{3'}$ during nucleophilic attack at P_α in the case that $O^{3'}$ is still protonated.

Later, Zhang and Salahub[17] performed DFT (PBE/DZVP-GGA) calculations on 91/94 atom models and favored a pathway with indirect proton transfer to $O^{P\alpha}$ via a water molecule, followed by proton transfer to $O^{P\beta}$.

Ramos et al.[18] investigated four possible pathways with different (initial) proton acceptors: $O^{P\alpha}$, OH^-, OH^- coordinated to Mg^{2+} and Asp485. The leaving pyrophosphate is protonated by a charged His1085. Geometries have been optimized at the ONIOM (B3LYP/6-31G(d):PM3MM) level, whereas free energies have been calculated at the B3LYP/6-31++G(2d,2p):PM3MM level of theory on models with up to 228 QM atoms. They found the hydroxyl ion to be the most likely proton acceptor and employed thermodynamic integration to verify the feasibility of the transfer from bulk solvent into the active site.

In the most recent work, Salahub et al.[19] considered three (initial) proton acceptors, $O^{P\alpha}$, Asp483, and H_2O coordinated to Mg^{2+}, and four different starting structures. 2D relaxed surface scans have been performed to include the possibility of a concerted mechanism. To allow the computation of 2D potential energy maps, the semiempirical AM1/d-PhoT[20] method was used, calibrated against B3LYP/TZVP. The authors found high barriers when starting from the crystal structures and significantly lower barriers when the starting structure is taken from MD simulation (“hybrid model”). They conclude that the crystal structures are outside of the catalytic pathway and favor direct or indirect protonation of $O^{P\alpha}$ followed by the $O^{3'}$ attack on P^α, proton migration to $O^{\alpha\beta}$, and pyrophosphate release.

In contrast to these computational studies and instead of investigating the overall fidelity of polymerases,[7,8,14,21] our present study focuses on the discrimination against 2′dNTPs in the chemical step. Although this is not the only source of discrimination, there are strong indications that it is a major source.[3,7,13,14] We explore the main difference between NTP and dNTP incorporated into the active site and the impact on the reaction barrier. In this work, we employ theoretical methods, including linear scaling quantum-chemical methods[22−25] within a QM/MM[26,27] approach, to provide deeper insights into the discrimination process. In order to understand the discrimination within the active site, we choose an X-ray structure that shows all contacts to the $O^{2'}$ mentioned above and create two systems containing NTP and dNTP, respectively. We use MD simulations of the full systems in explicit solvent (100 ns per system) to verify these contacts. We calculate reaction profiles of the nucleotide addition reaction on the DFT level, by geometry optimizations restraining the $O^{3'}-P^\alpha$ distance. These reaction profiles are further refined employing the nudged elastic band method, which avoids specification of a rigid reaction coordinate and samples the reaction coordinate equally. Since earlier studies have shown that often large QM spheres are necessary for a reliable theoretical description,[28−31] we converge the QM-size and find that 800 QM atoms are needed to consider all important residues. The reaction profiles are then recalculated with 800 QM atoms using linear scaling SCF methods.[22−25]

■ COMPUTATIONAL METHODS AND DETAILS

Molecular Dynamics. Force field MD simulations were performed with the NAMD 2.10 package[32] and the AMBER 99SB force field. Parameters for ATP were taken from Meagher et al.[33] VMD[34] was used for visualization, analysis, and image generation. The X-ray structure used for this work has been published by Cheung et al.[9] (PDB code 4A3F). It is the structure of a Pol II-DNA complex representing a minimal initially transcribing complex. Besides the enzyme, it contains fragments of 11 nucleotides of nontemplate DNA, 20 nucleotides of template DNA, 6 nucleotides of RNA, and 1 nucleotide triphosphate (NTP) analogue. The nucleotide addition cycle has been stopped by α,β-methyleneadenosine 5′-triphosphate (AMPCPP) as NTP analogue, preventing pyrophosphate formation, and with 5-bromouracil as a marker in the DNA template strand. As parts of the crystal structure are not resolved, the residues 1082−1096 of Rpb1 and Mg(B) were modeled via homology modeling based on the crystal structure 2E2H. All other missing residues are on the surface of the enzyme and were modeled using swiss pdb viewer[35] v4.0.4. AMPCPP and bromouracil were replaced by ATP/dATP and uracil, respectively. The LEAP module of AmberTools11[36] was used to add protons, neutralize the system with Na+ ions, and solvate it in a box of TIP3P water, using a buffer of 10 Å around the solute. In total, the simulated systems consist of 543186 atoms (ATP system) and 543176 atoms (dATP system). All MD simulations employed the particle mesh Ewald method,[37] periodic boundary conditions, Langevin dynamics for temperature control, and the Langevin piston Nosé−Hoover method for pressure control[38,39] (1 atm). The SETTLE algorithm[40] has been used to allow a time step of 2 fs in equilibration and production. Before the simulations, the solvent was optimized with 10000 steps of conjugate gradient minimization, followed by another 10000 steps, with restraints of 1 kcal/mol/Å^2 on the solute. The systems were heated to 300 K, increasing the temperature by 1 K/100 steps, while keeping the restraints on the solute. Each system was equilibrated for 600 ps, with decreasing restraints. For proper statistics, and to obtain a statistically significant analysis, production runs of 100 ns (5 × 10 ns/1 × 50 ns) have been simulated per system.

Energy Profiles. The nucleotide addition reaction consists of two steps: deprotonation of RNA at $H^{O3'}$ and attack of $O^{3'}$ at P^α of NTP. We start with the deprotonated intermediate by deleting $H^{O3'}$ and then calculate the reprotonation by the nearest water, to generate a suitable starting structure containing OH^-, and the attack at P^α. These reaction paths are calculated first by an adiabatic mapping approach with the trivial reaction coordinates $r_1 = d(O^{3'}-H)$ and $r_2 = d(O^{3'}-P^\alpha)$. The reaction path is then refined with the nudged elastic band (NEB) method[41−43] implemented in ChemShell,[44−47] using the resulting end structures and the deprotonated intermediate as initial guess. The advantages of the NEB method are that a minimal energy path (MEP) is found, connecting start and end structure. It is not necessary to specify the reaction coordinate with respect to bond distances, angles, or dihedrals. The resulting MEP will include the necessary structural changes. In addition, the NEB prevents hopping between different MEPs and samples the reaction coordinate equally, in contrast to the adiabatic mapping approach. Free QM/MM energies are calculated by the free energy perturbation (FEP) approach implemented in ChemShell,[48] including zero-point, thermal,

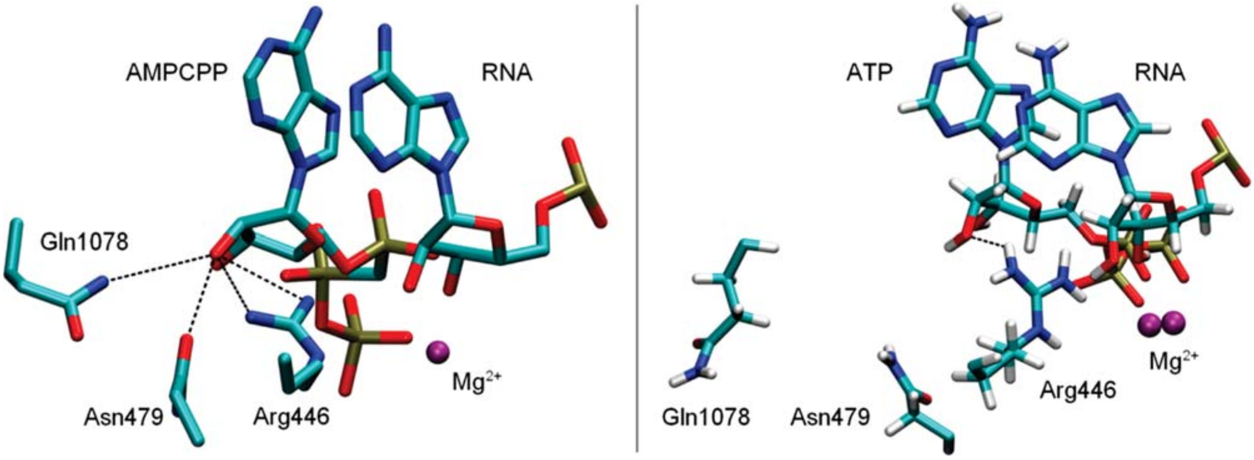


Figure 1. Contacts to $O^{2'}$ in the crystal structure 4A3F (left) and after MD (right).

and entropic corrections based on numeric frequencies for educt, transition state, and product.

QM/MM Model. The QM sphere includes all residues that might be directly involved in the mechanism (Figure 2): ATP, the terminal nucleotide of the RNA, OH^-, both Mg^{2+}, the coordination spheres of Mg^{2+} (all Asp and water within 4 Å), Arg446, His1085, and all water within 4 Å of His1085 and ATP. The model consists of 211 QM atoms. The total system has been reduced to all residues within a sphere of 30 Å around ATP, based on the influence on the interaction energy between ATP and Arg446 (Table S1). During geometry optimizations, all residues within 7 Å around the QM region are optimized, and the remaining MM region is fixed in position. For QM/MM structure optimizations the DL-POLY implementation within ChemShell[42] (AMBER-FF) was combined with density functional theory at the B3LYP-D3/6-31G** level of theory, employing the Q-Chem program package[49] for the QM part. We chose B3LYP-D3 due to its numeric stability and successful application on this particular system.[18,19]

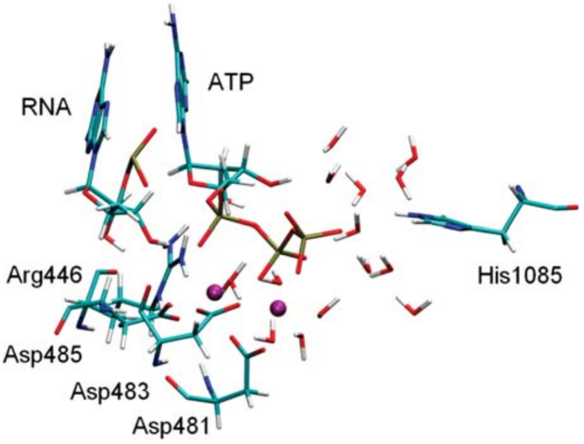


Figure 2. Smallest QM region employed (ATP) − for QM-size convergence see Figure 6.

■ RESULTS AND DISCUSSION

Contacts to $O^{2'}$. In the following, we show how the enzyme recognizes the sugar moiety and discriminates against 2′-deoxyribose in the nucleotide addition reaction. Most of the available crystal structures of RNA Pol II in complex with RNA and DNA show only one contact between the enzyme and the $OH^{2'}$ group of the NTP, which cannot account for the high selectivity observed experimentally.[3] However, Cramer et al.[9] found three amino acids (Arg446, Gln1078, Asn479) interacting with $OH^{2'}$ of the NTP (Figure 1). Therefore, we chose the corresponding structure with PDB code 4A3F as a starting structure. Unfortunately, we found that the distances between Gln1078/Asn479 and $OH^{2'}$ increase quickly during the first 0.5 ns of our MD simulations, and no H-bonds are formed within 50 ns of MD in both systems (ATP/dATP). This behavior could not be changed by additional restraints during the equilibration. The initial H-bonds to NTP break as soon as the restraints are reduced. Therefore, there is only the H-bond between $O^{2'}$-Arg446 in the case of ATP and $O^{3'}$-Arg446 in the case of dATP (Figure 5). In line with the induced fit discrimination observed experimentally,[7,50] we find much higher fluctuation in the dATP system compared to the ATP system (Table S2). Even base stacking between RNA and dATP is affected (Figure S1).

Nucleotide Addition Reaction. In the nucleotide addition reaction, $O^{3'}$ of the terminal RNA nucleotide attacks the NTP at P^{α}, elongating the RNA strand and releasing pyrophosphate. The exact mechanism is still under discussion, and based on computational investigations two different mechanisms come into question. Ramos et al.[18] favor deprotonation of the OH_{RNA} group by a bulk solvent hydroxide ion and protonation of O^{β}_{NTP} by His1085 followed by nucleophilic attack of the triphosphate by O^-_{RNA}, resulting in pyrophosphate cleavage. Salahub et al.[19] favor a direct or indirect proton transfer of $H^{O3'}_{RNA}$ to O^{α}_{NTP}, nucleophilic attack of the triphosphate by O^-_{RNA}, and proton transfer from O^{α}_{NTP} to $O^{\alpha\beta}_{NTP}$, followed by elimination of pyrophosphate.

We first applied the reaction mechanism suggested by Salahub et al.[19] to our setup and calculated the proton transfers of $H^{O3'}_{RNA}$ to O^{α}_{NTP} and of $H^{O3'}_{RNA}$ to both O^{D1} and O^{D2} of Asp483. None of these lead to a stable intermediate (Figures S2, S4). The combined reactions of proton transfer and

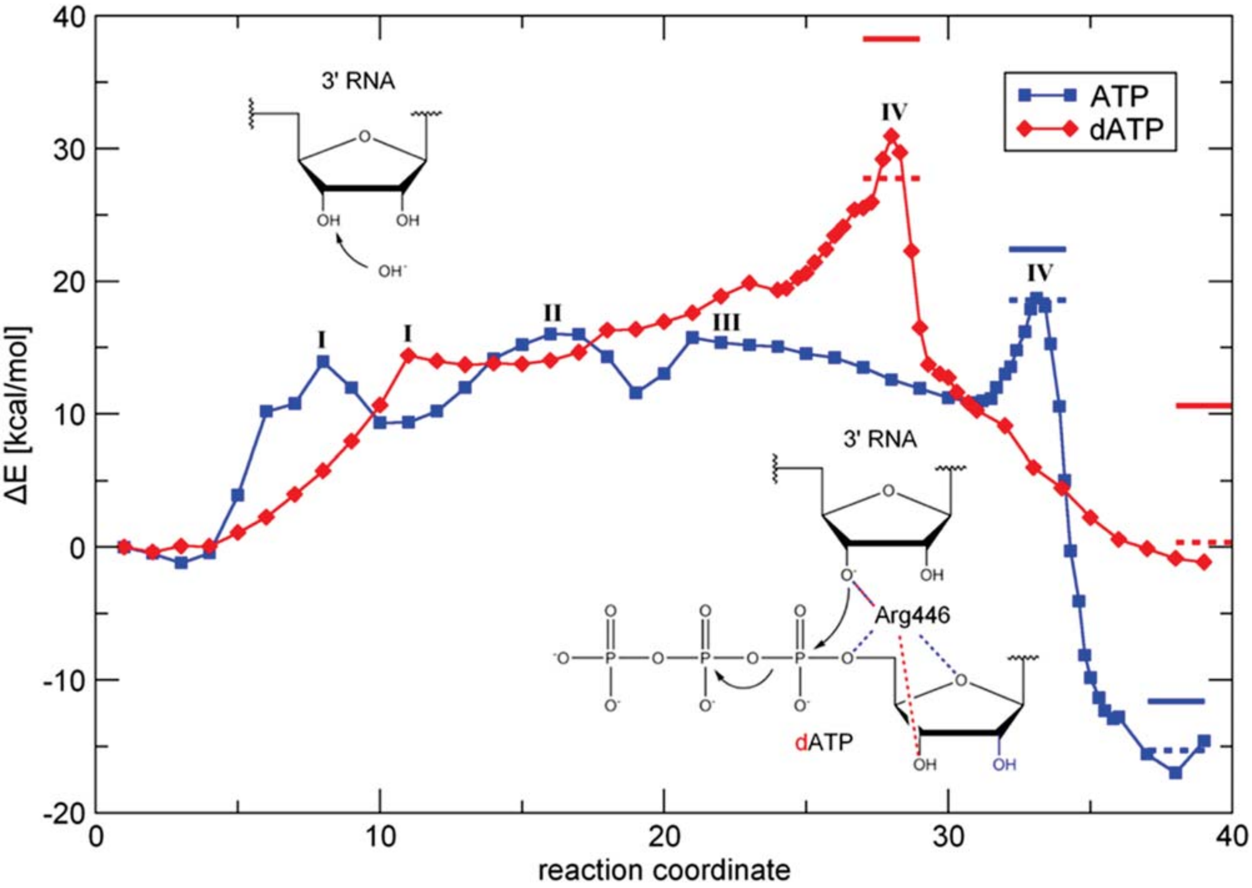


Figure 3. NEBs with 40 images based on adiabatic mapping reaction paths. The final barrier has been refined with 20 additional images. **I**: nucleophilic attack of OH^-, **II**: rearrangement of H_2O, **III**: movement of Arg446, **IV**: nucleophilic attack of $O^{3'-}_{RNA}$. In case of dATP, steps II and III do not occur due to the different orientation of Arg446. The dashed horizontal bars show the free QM/MM energy (FEP, the same QM region as the NEB), and the solid horizontal bars show the energy for converged QM size (800 QM atoms).

nucleotide addition lead to an unfavorable product, which is ~30 kcal/mol higher in energy than the educt (Figure S3, S5).

Instead, we will focus on the suggestion by Ramos et al.[18] and choose OH^- as the initial base, which leads to a stable intermediate with an energy ~10 kcal/mol above the educt. Subsequent nucleotide addition results in a stable product ~16 kcal/mol below the educt.

With the latter mechanism, we go beyond these studies and investigate the differences between the addition of ATP and dATP. The nucleotide addition reaction for both systems, calculated with the adiabatic mapping approach, is then further refined by the nudged elastic band (NEB) method. We find striking differences between the energy profiles of the reactions of ATP and dATP (Figure 3): After the initial deprotonation of $O^{3'}_{RNA}$ (I) with a barrier of 14 kcal/mol, the resulting water molecule rearranges (II) to allow Arg446 to bridge $O^{3'}_{RNA}$ and $O^{4'}_{ATP}$ in the subsequent step (III). Thus, the final barrier of the nucleophilic attack (IV) is only 8 kcal/mol, resulting in a total barrier of 19 kcal/mol. This is in excellent agreement with the estimated barrier of 18.1 kcal/mol based on the turnover number of the enzyme.[18]

In the last step, 36 kcal/mol are gained due to the simultaneous cleavage of the triphosphate, leading to a stable product 17 kcal/mol below the educt. In the case of dATP, the initial deprotonation of $O^{3'}_{RNA}$ (I) also requires 14 kcal/mol, but due to the different orientation of Arg446, rearrangement of the resulting water is not necessary. Therefore, the reaction coordinate continues with the nucleophilic attack (IV), resulting in a total barrier of 31 kcal/mol. By cleavage of the triphosphate, 32 kcal/mol are gained, resulting in a product with about the same energy as the educt. In contrast to ATP, the reaction of dATP has a considerably higher barrier and would be fully reversible, whereas the reaction of ATP shows a small energy barrier in perfect agreement with the experiment and leads to a stable product.

The calculation of free energies by a QM/MM-FEP approach (Figure S8, Table S3) does not change the picture significantly (Figure 3).

Another way to verify the results is to systematically increase the QM region until size-convergence is reached (details see below). The result emphasizes the difference between the reactions with ATP/dATP even more clearly. The energies for the converged QM region are marked by horizontal lines in Figure 3 and show that the barrier of the reaction with dATP increases to 38 kcal/mol and the product increases to 11 kcal/mol above the educt. The reaction with ATP still has a much lower barrier of 22 kcal/mol, and the product is 12 kcal/mol below the educt.

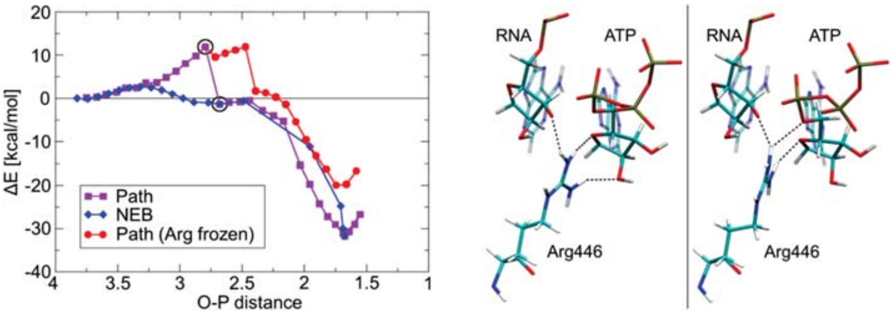


Figure 4. Influence of Arg446 movement on the energy profile. Left: energy profile of the nucleophilic attack of $O^{3'}$-RNA at P^{α}_{ATP} including 211 QM atoms. The energy gained between the two circled points of the adiabatic mapping path (magenta) corresponds to reorientation of Arg446. If the position of Arg446 is frozen at the first circled point, also the product energy is shifted by the amount of energy that would be gained due to arginine reorientation (red path). If the movement of Arg446 is included into the reaction coordinate of a NEB, the resulting barrier is reduced (NEB, blue). Right: The structures corresponding to the highlighted points in the energy profile. Rearrangement of Arg446 leads to a different H-bond pattern and results in a sudden gain of energy.

Influence of Arg446. Although the significance of Arg446 for the discrimination in RNA Pol II is well-known, its role could not be investigated experimentally, as mutation is lethal.[3,4] Therefore, we performed QM/MM calculations to elucidate the influence of Arg446 on the reaction profile and to reveal its role in the mechanism. Before the reaction, Arg446 forms H-bonds to $O^{3'-}$ of the terminal RNA and $O^{2'}$ and $O^{4'}$ of ATP. During the reaction the H-bond to $O^{2'}$ is broken and an H-bond to $O^{5'}$ is formed. Thereby ATP is slightly rotated around its long axis, the bridging by Arg446 is enhanced, and the formation of a bond between $O^{3'-}_{RNA}$ and P^{α} is facilitated. The impact of this effect can be estimated by an adiabatic mapping path, decreasing the $O^{3'}-P^{\alpha}$ distance as the reaction coordinate, starting from deprotonated RNA. The sudden gain of 13 kcal/mol corresponds to the movement of Arg446. An NEB is able to include this movement into the reaction coordinate to find a reaction path with a significantly lower barrier (Figure 4). In the case of dATP, Arg446 forms an H-bond to $O^{3'}$. Thereby, dATP is shifted in comparison to ATP. The initial distance between $O^{3'}_{RNA}$ and P^{α} of dATP is increased by 1 Å, making it more difficult to form a bond (Figure 5). This also means that Arg446 cannot connect dATP and the terminal RNA to facilitate bond formation.

Influence of the QM Region. The choice of the QM region can have a huge impact on the results. Especially small QM sizes have been shown to be problematic and can easily be misleading.[29−31,51] It has been shown that the QM size, necessary to include all important residues, can be determined by single-point calculations with successively increasing QM regions.[31,52] Therefore, we increase the QM region and include all residues that have at least one atom within a certain distance to the initial QM region. Although single point calculations are sufficient to identify the converged QM size, geometry optimizations are necessary to determine the actual energy difference. In analogy to recent work,[52,53] the active region has been limited to the previous QM region, preventing structural changes of the environment unrelated to the reaction mechanism. Educt and product geometries are optimized without restraints, whereas the geometries around the highest barrier are optimized restraining the reaction coordinate (Figures S6, S7). We find that the QM size is converged with respect to the energy at about 800 QM atoms, see Figure 6. The solid horizontal bars in Figure 3 show the converged energies within the reaction profile.

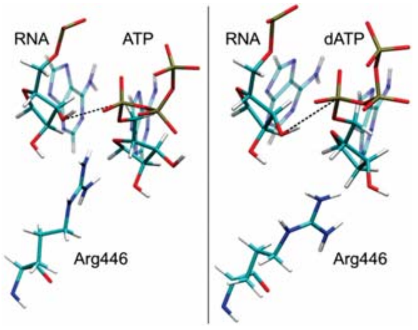


Figure 5. Relative position of Arg446, NTP (left: ATP, right dATP), and 3′ terminal RNA after MD simulation.

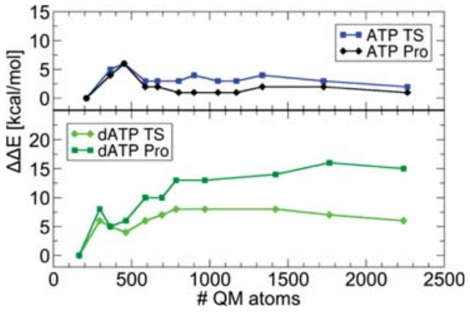


Figure 6. QM size convergence of the energy difference between educt and transition state (TS, light color) and product (Pro, dark color). ΔE of the smallest QM size is set to zero.

■ CONCLUSIONS

When RNA polymerase II catalyzes the polymerization of RNA from nucleoside triphosphates, it not only selects NTPs with a base complementary to the DNA template but also discriminates against 2′dNTPs with an error rate of 10^{-5} per nucleotide.[1] Kinetic data show that this discrimination takes place mainly in the reaction step,[3,7] and X-ray structures revealed up to three H-bonds to $O^{2'}$ that might be responsible. Starting from the X-ray structure containing the three H-bonds,[9] our FF-MD simulations indicate that only the hydrogen bond to Arg446 is stable. We showed that this interaction is necessary to reach a catalytically active conformation, resulting in a discrimination energy of 16 kcal/mol for QM size-converged barriers (9 kcal/mol for free energy barriers with smaller QM region). Using a nudged elastic band approach, we find a stepwise mechanism with OH^- as the initial base, contrasting recent computational work that employed a different X-ray structure.[18] For ATP, our calculated reaction barrier of 22 kcal/mol (18 kcal/mol, free energy) is in excellent agreement with the estimated barrier of 18.1 kcal/mol,[18] that is based on the experimental turnover number of the enzyme. The large influence of the orientation of Arg446, resulting in a 13 kcal/mol higher reaction barrier, finally explains why all mutations were found to be lethal[3,4] and how a single H-bond can account for the observed discrimination. We hope that these findings will also help to improve the understanding of the mechanistic details of ribose discrimination in other polymerases.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00157.

Additional computational details (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: christian.ochsenfeld@uni-muenchen.de.

ORCID

Christian Ochsenfeld: 0000-0002-4189-6558

Funding

We acknowledge financial support by the Volkswagen Stiftung within the funding initiative “New Conceptual Approaches to Modeling and Simulation of Complex Systems” and the DFG cluster of excellence EXC 114 “Center for Integrative Protein Science Munich” (CIPSM). In addition, some computational time was provided by the LRZ Munich.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Professor Patrick Cramer (MPI Goettingen) for useful discussions on RNA polymerase.

■ REFERENCES

(1) Ninio, J. Connections between Translation, Transcription and Replication Error-Rates. *Biochimie* **1991**, *73* (12), 1517−1523.

(2) Svetlov, V.; Vassylyev, D. G.; Artsimovitch, I. Discrimination against Deoxyribonucleotide Substrates by Bacterial RNA Polymerase. *J. Biol. Chem.* **2004**, *279* (37), 38087−38090.

(3) Wang, D.; Bushnell, D. A.; Westover, K. D.; Kaplan, C. D.; Kornberg, R. D. Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* **2006**, *127* (5), 941−954.

(4) Kaplan, C. D.; Jin, H.; Zhang, I. L.; Belyanin, A. Dissection of Pol II Trigger Loop Function and Pol II Activity-Dependent Control of Start Site Selection in Vivo. *PLoS Genet.* **2012**, *8* (4), e1002627.

(5) Huang, X.; Wang, D.; Weiss, D. R.; Bushnell, D. A.; Kornberg, R. D.; Levitt, M. RNA Polymerase II Trigger Loop Residues Stabilize and Position the Incoming Nucleotide Triphosphate in Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (36), 15745−15750.

(6) Castro, C.; Smidansky, E.; Maksimchuk, K. R.; Arnold, J. J.; Korneeva, V. S.; Götte, M.; Konigsberg, W.; Cameron, C. E. Two Proton Transfers in the Transition State for Nucleotidyl Transfer Catalyzed by RNA- and DNA-Dependent RNA and DNA Polymerases. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (11), 4267−4272.

(7) Yuzenkova, Y.; Bochkareva, A.; Tadigotla, V. R.; Roghanian, M.; Zorov, S.; Severinov, K.; Zenkin, N. Stepwise Mechanism for Transcription Fidelity. *BMC Biol.* **2010**, *8* (1), 54.

(8) Kellinger, M. W.; Ulrich, S.; Chong, J.; Kool, E. T.; Wang, D. Dissecting Chemical Interactions Governing RNA Polymerase II Transcriptional Fidelity. *J. Am. Chem. Soc.* **2012**, *134* (19), 8231−8240.

(9) Cheung, A. C. M.; Sainsbury, S.; Cramer, P. Structural Basis of Initial RNA Polymerase II Transcription. *EMBO J.* **2011**, *30* (23), 4755−4763.

(10) Gnatt, A. L.; Cramer, P.; Fu, J.; Bushnell, D. A.; Kornberg, R. D. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. *Science* **2001**, *292* (5523), 1876−1882.

(11) Vassylyev, D. G.; Vassylyeva, M. N.; Zhang, J.; Palangat, M.; Artsimovitch, I.; Landick, R. Structural Basis for Substrate Loading in Bacterial RNA Polymerase. *Nature* **2007**, *448* (7150), 163−168.

(12) Zhang, J.; Palangat, M.; Landick, R. Role of the RNA Polymerase Trigger Loop in Catalysis and Pausing. *Nat. Struct. Mol. Biol.* **2010**, *17* (1), 99−104.

(13) Zhu, R.; de la Lande, A.; Zhang, R.; Salahub, D. R. Exploring the Molecular Origin of the High Selectivity of Multisubunit RNA Polymerases by Stochastic Kinetic Models. *Interdiscip. Sci.: Comput. Life Sci.* **2009**, *1* (2), 91−98.

(14) Wang, B.; Opron, K.; Burton, Z. F.; Cukier, R. I.; Feig, M. Five Checkpoints Maintaining the Fidelity of Transcription by RNA Polymerases in Structural and Energetic Details. *Nucleic Acids Res.* **2015**, *43* (2), 1133−1146.

(15) Beese, L. S.; Steitz, T. A. Structural Basis for the 3′-5′ Exonuclease Activity of Escherichia Coli DNA Polymerase I: A Two Metal Ion Mechanism. *EMBO J.* **1991**, *10* (1), 25−33.

(16) Zhu, R.; Salahub, D. R.; Maroulis, G.; Simos, T. E. Mechanisms of Nucleotidyl Transfer Catalyzed by the Yeast RNA Polymerase II. *AIP Conf. Proc.* **2007**, *963*, 104−110.

(17) Zhang, Y.; Salahub, D. A Theoretical Study of the Mechanism of the Nucleotidyl Transfer Reaction Catalyzed by Yeast RNA Polymerase II. *Sci. China: Chem.* **2012**, *55* (9), 1887−1894.

(18) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. The Catalytic Mechanism of RNA Polymerase II. *J. Chem. Theory Comput.* **2011**, *7* (4), 1177−1188.

(19) Zhang, R.; Bhattacharjee, A.; Field, M. J.; Salahub, D. R. Multiple Proton Relay Routes in the Reaction Mechanism of RNAP II: Assessing the Effect of Structural Model. *Proteins: Struct., Funct., Genet.* **2015**, *83* (2), 268−281.

(20) Nam, K.; Gao, J.; York, D. M. Quantum Mechanical/Molecular Mechanical Simulation Study of the Mechanism of Hairpin Ribozyme Catalysis. *J. Am. Chem. Soc.* **2008**, *130* (14), 4680−4691.

(21) Florián, J.; Goodman, M. F.; Warshel, A. Computer Simulations of Protein Functions: Searching for the Molecular Origin of the Replication Fidelity of DNA Polymerases. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6819−6824.

(22) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. The Continuous Fast Multipole Method. *Chem. Phys. Lett.* **1994**, *230* (1−2), 8−16.

(23) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and Sublinear Scaling Formation of Hartree-Fock-Type Exchange Matrices. *J. Chem. Phys.* **1998**, *109* (5), 1663−1669.

(24) Ochsenfeld, C. Linear Scaling Exchange Gradients for Hartree−Fock and Hybrid Density Functional Theory. *Chem. Phys. Lett.* **2000**, *327*, 216−223.

(25) Kussmann, J.; Beer, M.; Ochsenfeld, C. Linear-Scaling Self-Consistent Field Methods for Large Molecules. *WIREs Comput. Mol. Sci.* **2013**, *3* (6), 614−636.

(26) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48* (7), 1198−1229.

(27) Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53* (38), 10020−10031.

(28) Sumowski, C. V.; Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **2009**, *113* (43), 11734−11741.

(29) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8* (10), 3793−3803.

(30) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8* (7), 2260−2271.

(31) Sumner, S.; Söderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4205−4214.

(32) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.

(33) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of Polyphosphate Parameters for Use with the AMBER Force Field. *J. Comput. Chem.* **2003**, *24* (9), 1016−1025.

(34) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.

(35) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An Environment for Comparative Protein Modeling. *Electrophoresis* **1997**, *18* (15), 2714−2723.

(36) Case, D. A.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *Amber 11*; University of California: San Francisco, CA, 2010.

(37) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New Tricks for Modelers from the Crystallography Toolkit: The Particle Mesh Ewald Algorithm and Its Use in Nucleic Acid Simulations. *Structure* **1999**, *7* (3), R55−R60.

(38) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177.

(39) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613.

(40) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952−962.

(41) Jonsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific; 1998; pp 385−404.

(42) Henkelman, G.; Jonsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113* (22), 9978−9985.

(43) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113* (22), 9901−9904.

(44) ChemShell, a Computational Chemistry Shell. See www.chemshell.org (accessed Feb 28, 2017).

(45) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; et al. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1−28.

(46) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113* (43), 11856−11865.

(47) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell - a Modular Software Package for QM/MM Simulations. *WIREs Comput. Mol. Sci.* **2014**, *4* (2), 101−110.

(48) Kästner, J.; Senn, H. H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction. *J. Chem. Theory Comput.* **2006**, *2* (2), 452−461.

(49) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184−215.

(50) Seibold, S. A.; Singh, B. N.; Zhang, C.; Kireeva, M.; Domecq, C.; Bouchard, A.; Nazione, A. M.; Feig, M.; Cukier, R. I.; Coulombe, B.; et al. Conformational Coupling, Bridge Helix Dynamics and Active Site Dehydration in Catalysis by RNA Polymerase. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2010**, *1799* (8), 575−587.

(51) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do Quantum Mechanical Energies Calculated for Small Models of Protein-Active Sites Converge? *J. Phys. Chem. A* **2009**, *113* (43), 11793−11800.

(52) Blank, I. D.; Sadeghian, K.; Ochsenfeld, C. A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination Within the Active Site. *Sci. Rep.* **2015**, *5*, 10369.

(53) Sadeghian, K.; Flaig, D.; Blank, I. D.; Schneider, S.; Strasser, R.; Stathis, D.; Winnacker, M.; Carell, T.; Ochsenfeld, C. Ribose-Protonated DNA Base Excision Repair: A Combined Theoretical and Experimental Study. *Angew. Chem., Int. Ed.* **2014**, *53*, 10044−10048.

Supporting Information for

A quantum-chemical study of the discrimination against dNTP in the nucleotide addition reaction of RNA Polymerase II

Sven Roßbach[†, ‡] and Christian Ochsenfeld[†, ‡,*]

[†] Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany

[‡] Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr, 5-13, D-81377 Munich, Germany

1. Size of the system

To estimate the size of the system, we calculated interaction energies between the crucial residue Arg446 and ATP for different cutouts. We conclude that a sphere with a radius of 30Å is sufficient to take the enzyme into account properly (Table S1).

Table S1: Interaction energy between Arg446 and ATP depending on the radius of the MM sphere.

radius	interaction energy
20 Å	-133.8 kcal/mol
30 Å	-139.4 kcal/mol
50 Å	-139.3 kcal/mol

2. Root mean square fluctuation (RMSF)

The comparison of the RMSF over all our MD simulations (Table S2) shows clearly, that there is significantly less fluctuation in the active site with a matching ATP than with a non-matching dATP. This is in line with an induced fit mechanism, facilitating the reaction only in case of a matching substrate.

Table S2: RMSF of active site residues over 100 ns of MD in Å.

residue	ATP	dATP
Arg446	0.91	1.48
Asn479	1.22	1.81
Gln1078	1.61	2.27
Leu1081	2.29	2.61
Asn1082	2.59	2.46
His1085	2.49	2.64
Arg766	1.13	1.41
Lys987	0.91	1.36
Arg1020	1.35	1.65
ATP/dATP	0.89	1.69

3. Base-stacking (50 ns runs)

The base-stacking between dATP and the 3'-terminal RNA base is regularly lost and reformed during 50 ns of MD, while in case of ATP, the same interaction is stable all the time (Figure S1).

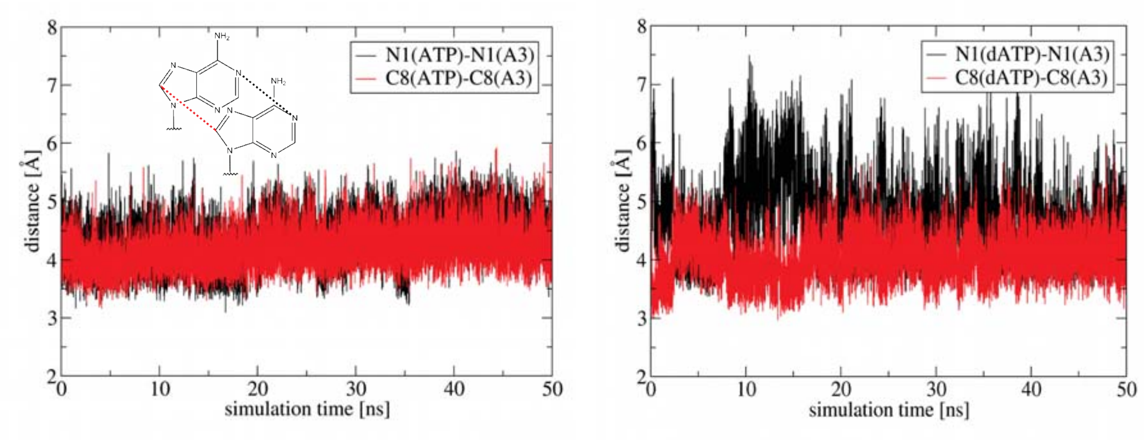


Figure S1: Distance between the N1 atoms / C8 atoms of ATP/dATP and the 3' terminal RNA base.

4. Reaction via Asp485

The first step in the nucleotide addition reaction (deprotonation of $O^{3'}_{RNA}$) by Asp485 does not lead to a minimum structure (Figure S2), therefore Asp485 is not a suitable proton acceptor in this case. Also the concerted reaction with simultaneous pyrophosphate release does not lead to a favorable product (Figure S3).

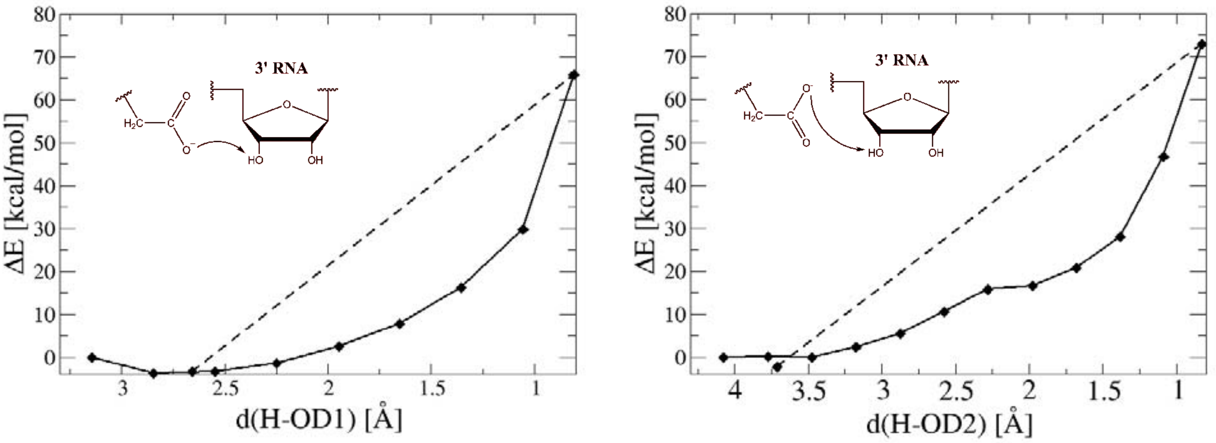


Figure S2: Proton transfer $H^{O3'}_{RNA} - O^{D1}_{Asp485}$ (left), proton transfer $H^{O3'}_{RNA} - O^{D2}_{Asp485}$ (left). The dashed line means free optimization of the last geometry.

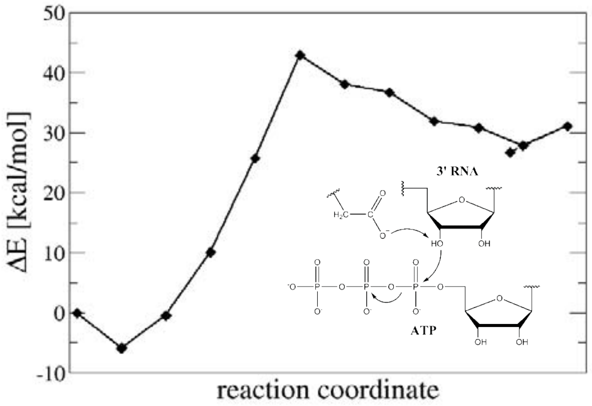


Figure S3: Reaction profile of the nucleotide addition reaction including proton transfer to O^{D1}_{Asp485}.

5. Reaction via ATP:

The deprotonation of $O^{3'}_{RNA}$ by ATP does not lead to a minimum structure (Figure S4) and also the concerted reaction with addition of ATP and cleavage of pyrophosphate has a high barrier and does not lead to a favorable product (Figure S5). Therefore, O^{1A}_{ATP} is not a suitable proton acceptor in this case.

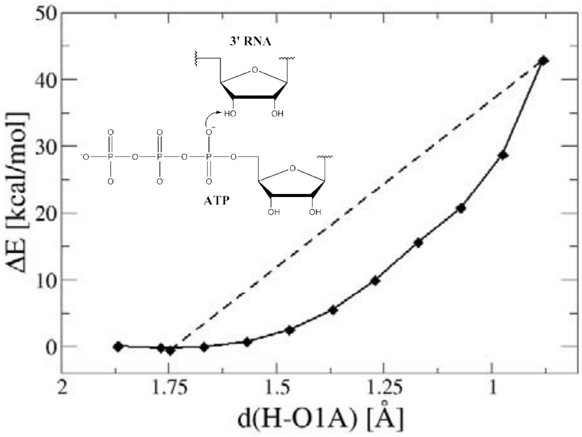


Figure S4: Proton transfer $H^{O3'}_{RNA} – O^{1A}_{ATP}$.
The dashed line means optimization without restraints.

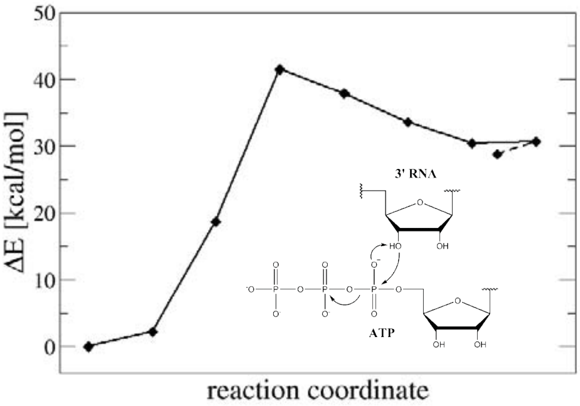


Figure S5: Reaction profile of the nucleotide addition reaction including proton transfer to O^{1A}_{ATP}.

6. Size converged reaction barriers

The geometries around the highest barrier have been re-optimized for both setups (ATP/dATP) with the converged QM size, restraining the reaction coordinate and the surroundings to prevent unrelated geometrical changes. In both cases, the relative position of the maximum on the reaction coordinate does not change. With the converged QM size, the difference between the reaction barrier of ATP and the reaction barrier of dATP increases even further (Figure S6, S7).

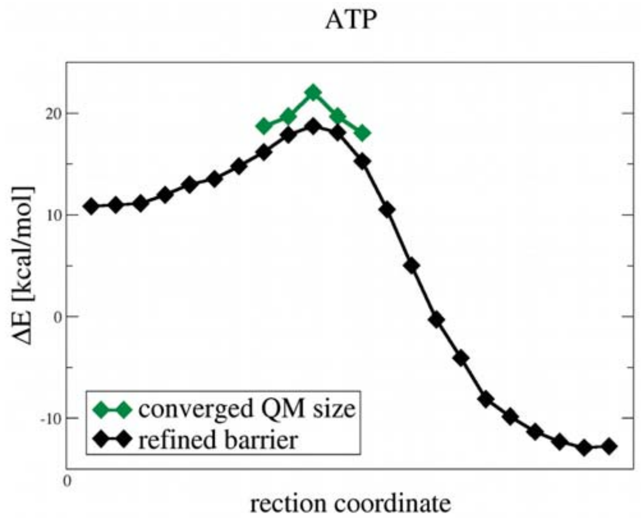


Figure S6: Nucleophilic attack of O^{3-}_{RNA}, refined with an additional NEB (black) and restrained optimizations with 800 QM atoms (green).

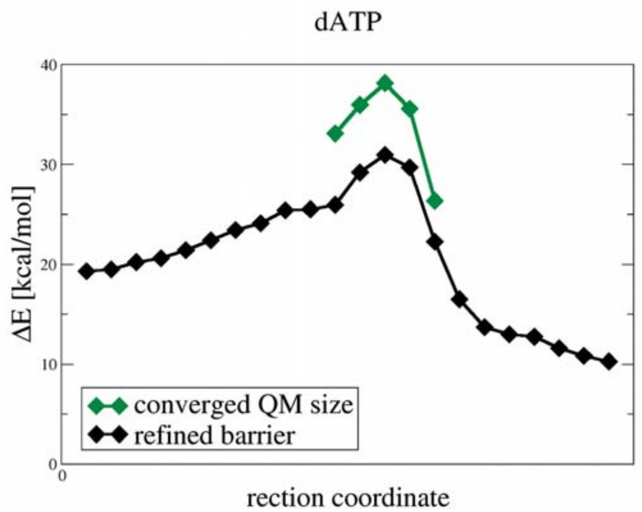


Figure S7: Nucleophilic attack of O^{3-}_{RNA}, refined with an additional NEB (black) and restrained optimizations with 800 QM atoms (green).

7. Free energy calculations

The free energy differences have been calculated with the QM/MM free energy perturbation (FEP) method that includes the entropy contributions of the MM part and the QM/MM interaction (Figure S8). Additionally, the entropy of the QM part is included based on harmonic frequency calculations (Table S3). Although there is a stronger impact on the reaction profile of dATP than on the reaction profile of ATP, the reaction barriers do not change significantly.

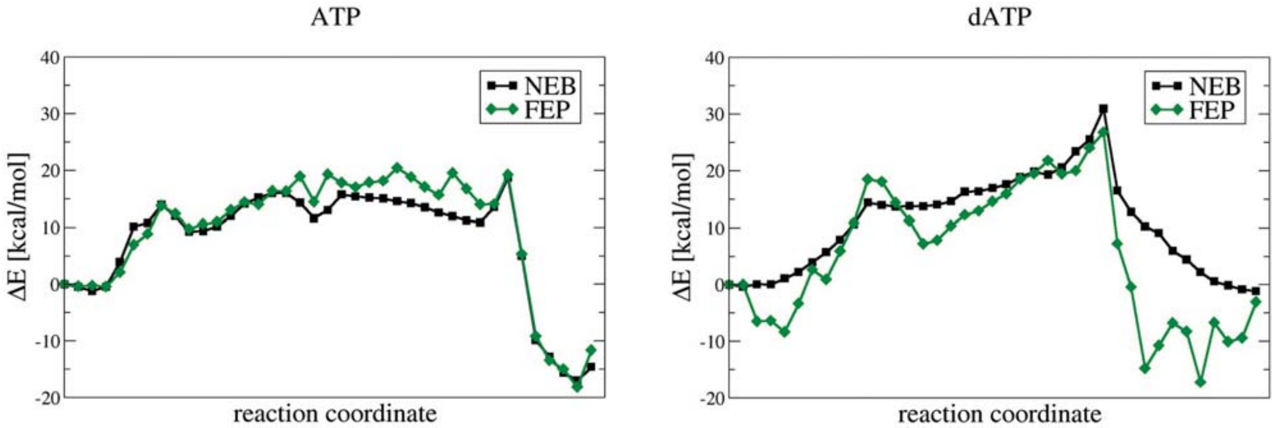


Figure S8: Reaction profile of the full nucleotide addition reaction with initial deprotonation of the 3'-RNA end and subsequent attack at P^{α}_{NTP}. The nudged elastic band profile, based on adiabatic mapping reaction paths, is shown in black. The free energy perturbation profile shown in green is the average of forward and backward perturbation.

Table S3: The correction of the QM part ($\Delta\Delta E_{QM}$) includes zero-point, thermal and entropic corrections based on numeric frequencies.

	ATP $\Delta\Delta E_{QM}$ (kcal/mol)	dATP $\Delta\Delta E_{QM}$ (kcal/mol)
transition state (final barrier)	-0.76	0.92
product	2.85	3.55

Paper 2: “Influence of coupling and embedding schemes on the QM size convergence in QM/MM approaches for the example of a proton transfer in DNA”, S. Roßbach, C. Ochsenfeld, *J. Chem. Theory Comput.*, **2017**, *13*, 1102-1107.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Influence of Coupling and Embedding Schemes on QM Size Convergence in QM/MM Approaches for the Example of a Proton Transfer in DNA

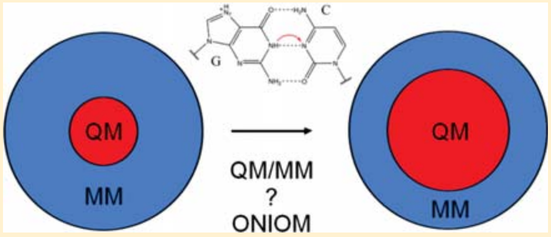
Sven Roßbach[†,‡] and Christian Ochsenfeld*[,†,‡]

[†] Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 7, D-81377 Munich, Germany

[‡] Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU Munich), Butenandtstr, 5-13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: The influence of embedding and coupling schemes on the convergence of the QM size in the QM/MM approach is investigated for the transfer of a proton in a DNA base pair. We find that the embedding scheme (mechanical or electrostatic) has a much greater impact on the convergence behavior than the coupling scheme (additive QM/MM or subtractive ONIOM). To achieve size convergence, QM regions with up to 6000 atoms are necessary for pure QM or mechanical embedding. In contrast, electrostatic embedding converges faster: for the example of the transfer of a proton between DNA base pairs, we recommend including at least five base pairs and 5 Å of solvent (including counterions) into the QM region, i.e., a total of 1150 atoms.



INTRODUCTION

Over the last 2 decades, the combined quantum mechanical/molecular mechanical (QM/MM) approach received increasing attention[1,2] as it became evident that small QM models in the gas phase are often insufficient to properly describe real systems. In particular, studies of large biomolecular systems and processes in explicit solvent often require models with thousands or even tens of thousands of atoms that cannot be treated quantum mechanically to date. The QM/MM approach usually leads to a better description of the system compared to a small pure QM model, as the environment is fully taken into account. Recently, it has been shown that large QM regions are also important within the QM/MM scheme and that energy differences converge much faster and are more stable with regard to the QM size when employing the QM/MM approach.[3−10]

There are two possible ways to describe the coupling between QM and MM: additive and subtractive (Figure 1). In the additive case, which we will abbreviate in the following as QM/MM, the energy is the sum of the MM energy (without all terms already described by QM), the QM energy, and the energy of an explicit coupling term.[1,11] In the subtractive case (ONIOM), the total energy consists of the MM energy of the full system, minus the MM energy of the QM region, plus the QM energy of the QM region.[12,13]

The (additive) QM/MM approach has the advantage that parameters for QM and link atoms, saturating covalent bonds between QM and MM, are unnecessary, as these are never described by the force field. The subtractive ONIOM approach requires accurate parameters for all atoms, including link atoms, because an MM calculation of the QM region is also necessary to avoid double counting. Without link atoms and with electrostatic embedding, dummy parameters could be used for the QM region. In the QM/MM approach, only one calculation is needed for each region, but the ONIOM approach allows for any number of layers, combining any methods, including only QM methods (QM/QM). While the QM/MM approach needs a special coupling term to treat the border, in the ONIOM approach, the coupling is described on the lower level with the aim that possible artifacts derived from link atoms are implicitly corrected if the force field can reproduce the QM energies for the link atoms.[14]

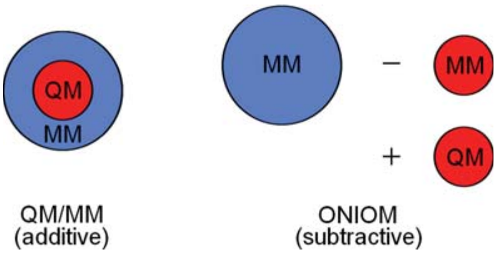


Figure 1. Schematic partitions by the additive QM/MM approach and the subtractive ONIOM method.

Received: July 21, 2016
Published: February 14, 2017

DOI: 10.1021/acs.jctc.6b00727
J. Chem. Theory Comput. 2017, 13, 1102−1107

Independent of the coupling scheme, the QM region can be embedded in the MM region in different ways. Mechanical embedding treats the QM−MM interactions at the MM level, neglecting the mutual polarization of both regions. Electrostatic embedding includes the MM point charges into the QM calculation, allowing polarization of the QM region. Independent of the embedding scheme, QM size convergence within a QM/MM approach has been studied for different systems employing only the additive coupling scheme[3−7,10] or only the subtractive coupling scheme.[8,15−18]

We are interested in the influence of the

- coupling scheme: subtractive or additive
- embedding scheme: mechanical or electrostatic

on the QM size convergence. It has been suggested that the charges of the atoms in the QM region should be fitted based on the corresponding QM calculation.[17] However, in this work, the original mechanical embedding[12,19] has been employed. For our investigation, we chose an extensively studied proton transfer reaction within a DNA base pair (Figure 2), which has been repeatedly shown to require a larger QM region than previously expected:

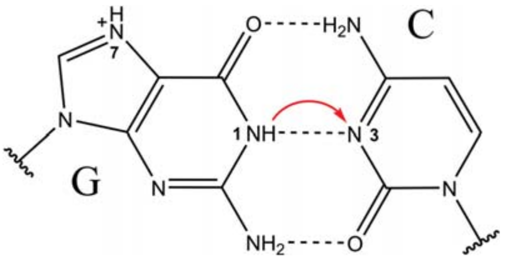


Figure 2. Single proton transfer within a protonated CG base pair, corresponding to residues 7 and 14 of the crystal structure.

In 1963, Löwdin[20] suggested "quantum jumps" (proton transfer) within base pairs as a source of mutations, which have been studied since then. Early calculations focused on the transfer of a single proton between the two bases in vacuum[21] or employed semiempirical methods to investigate double proton transfer.[22,23] Later, more accurate calculations employed perturbation theory or density-functional theory with increasing basis sets to study the stability of tautomers,[24−26] proton affinities of single bases,[25,27] and the dynamics of a base pair.[28−30] Nevertheless, single and double proton transfers between two base pairs have remained under investigation, with special attention given to the influence of a protonated base pair,[31,32] base stacking,[33−35] backbone and counterions,[33] and the solvent effect treated as a continuum model[24,25,28] or explicit microhydration.[26,33,36,37] It was suggested,[24] and later verified,[38] that the energetics of proton transfers also depend on the sequence. Therefore, very recent work added the adjacent base pairs to their models and included the solvent effect by explicit and/or implicit treatment of water.[39−41]

In our present work, we study systematically the QM size convergence of different embedding and coupling schemes for the example of the two DNA protonation states outlined above. Here, we investigate only the convergence behavior and not the absolute value of the energy difference itself.

COMPUTATIONAL DETAILS

Molecular Dynamics. The initial X-ray structure used for this work (1ZEW)[42] contains 10 base pairs of B-DNA. We used the LEAP module of AmberTools11[43] to add protons, neutralize the system with Na^+ ions, and solvate it in a box of TIP3P water with a buffer of 10 Å around the DNA, resulting in 13 127 atoms. Force field MD simulations were performed with the NAMD 2.10 package[44] and the AMBER10 force field. For the protonated nucleotides (Figure 2), GAFF parameters[45] were assigned with Antechamber.[46] Before the simulations, the solvent was optimized with 10 000 steps of conjugate gradient minimization, followed by another 10 000 steps, with restraints of 1 kcal/mol/$Å^2$ on the DNA. The system was heated to 300 K, increasing the temperature by 1 K/100 steps, while keeping the restraints. To allow a time step of 2 fs during equilibration, the SETTLE algorithm[47] was employed. Moreover, we used periodic boundary conditions, the particle mesh Ewald method,[48] Langevin dynamics for temperature control, and the Langevin piston Nosé−Hoover method for pressure control[49,50] (1 atm). The system was equilibrated for 400 ps, with restraints on the DNA, to obtain proper solvation. Residue 7 was then additionally protonated at N7, and the system was minimized again. To prepare the system after proton transfer, residue 7 was deprotonated at N1 and residue 14 was protonated at N3, followed by minimization of the bases.

QM Size Convergence. To study the convergence of the QM size, we calculated the energy difference of the two protonation states on the FF-minimized structures and systematically increased the QM region. We added base pairs until all 10 base pairs were included; then, we included water and ions based on the minimal distance of any atom to the DNA. Note that the absolute energy difference is not meaningful, as the structures are not reoptimized for any QM size. It has been shown that reoptimization of the geometries is not necessary to find the converged QM size[8,10] because single-point calculations follow the same trend and allow the calculation of larger QM regions. Moreover, as we model both protonation states, independent geometry optimizations would result in different conformations of the solvent and therefore lead to energy differences, which are mainly unrelated to proton transfer. Distance-based QM size convergence would thus become impossible. All QM calculations were performed with FermiONs++.[51,52]

ONIOM in ChemShell. In ChemShell,[11,53,54] the favored method to couple the QM and MM regions is by electrostatic embedding with the link atom approach in case there is a bond across the QM−MM border.[55] To avoid overpolarization of the link atom by the point charge of the neighboring MM atom, its charge is shifted to the next layer of MM atoms and the dipole moment is preserved by the addition of dipoles on these atoms.[11] To the "subtractive" module of ChemShell, we added the possibility of including electrostatic embedding with link atoms, shifted charges and dipoles in the MM calculation of the QM region. In this case, the dipoles are of the AMBER type EP and all residues containing link atoms need additional parameters, which are assigned by Antechamber[46] with RESP charges[56,57] at the B3LYP[58] level of theory with the 6-31G* basis set,[59,60] without structure reoptimization. The RESP charges were fitted on a grid of $N_{\text{atoms}} \times 2000$ points on the van der Waals shell scaled by 1.5.

■ RESULTS AND DISCUSSION

Figure 3 shows the energy difference between the two protonation states with an increasing number of base pairs in the QM region for all combinations of the coupling and embedding schemes. While our focus here is on the convergence behavior and not on the value of Δ*E* (see Computational Details, QM Size Convergence), the most apparent differences in Figure 3 are the slow convergence of the ME scheme and that it seems to converge to a different Δ*E* than EE. In contrast, we find that the coupling scheme has only a minor influence on QM size convergence and that the first layer of adjacent base pairs has a very significant impact on Δ*E*, whereas in the case of EE, the influence of all base pairs beyond the second layer can be sufficiently described by point charges.

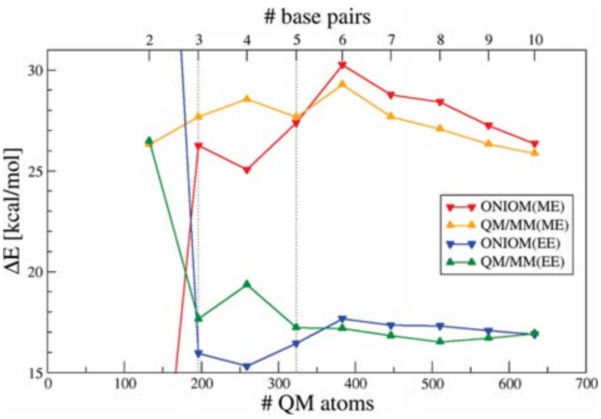


Figure 3. QM size convergence of additive (QM/MM) vs subtractive (ONIOM) coupling with mechanical (ME) and electrostatic (EE) embedding (M06-2X-D3/def2-SVP). The dashed lines show the first and second layers of base pairs around the proton transfer.

Further increasing the QM region to include solvent (water and Na^+ ions) eliminates the difference between the coupling schemes and shows that ME converges to the same Δ*E* as EE, but for ME, even 4000 QM atoms are not enough to reach size convergence (Figure 4).

Furthermore, no embedding scheme (pure QM) should also converge to the same energy difference as the QM−MM methods. Recent findings[3,7,8,61] showed that large energy fluctuations are usually expected when increasing the QM size in the gas phase before size convergence is reached. Typically, 1000 QM atoms are clearly sufficient, even though, depending on the accuracy, sometimes more than 1000 QM

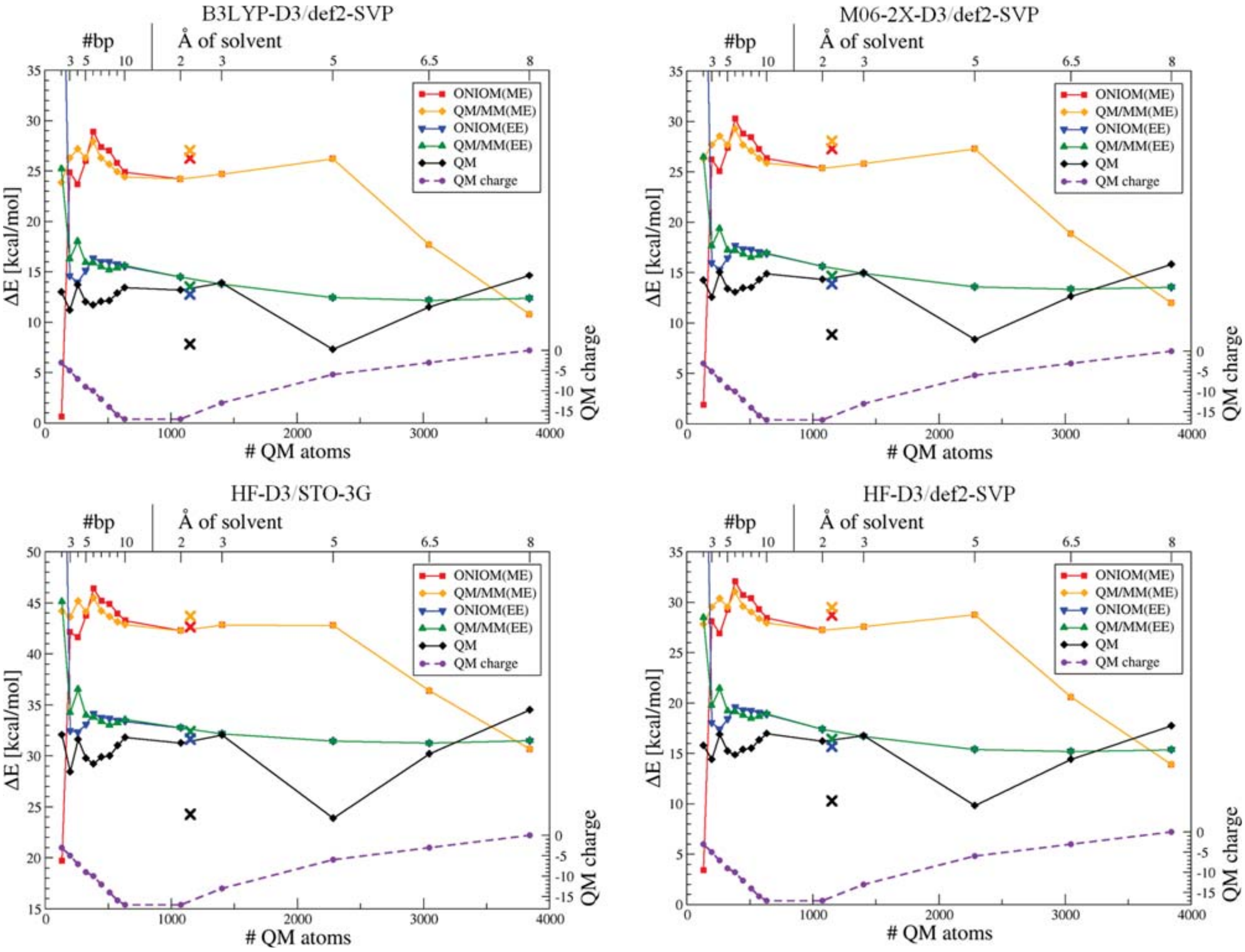


Figure 4. Comparison of QM size convergence for different method/basis set combinations. In addition, crosses show data for the QM region with five base pairs and 5 Å of solvent (water and ions).

atoms are needed. In our case, ΔE(pure QM) seems to converge rather fast and becomes stable with fluctuations of less than 2.6 kcal/mol within 1000 atoms with def2-SVP[62] (Figure 4). However, further increasing the QM region shows that the apparent convergence is misleading and that oscillations occur beyond 1000 atoms; even 4000 atoms are not enough to converge ΔE(pure QM) with respect to the QM size. It seems that the total QM charge can have a significant influence on ΔE, although changes in the total QM charge do not necessarily lead to a change in the energy difference.

We find that the convergence behavior of different methods is quite independent of the choice of method and basis set when comparing B3LYP-D3,[58,63] M06-2X-D3,[63,64] and HF-D3 with an SVP basis. Even HF-D3/STO-3G shows basically the same picture, which allows us to further extend the QM region. The energy difference for pure QM finally converges to the same value as that with EE when we include 6000 atoms in the QM region (Figure 5). Additionally, ME finally converges to the same ΔE as that of EE and pure QM, but even with 6000 QM atoms, corresponding to 10 base pairs including 11.3 Å of solvent, ΔE(ME) differs by 2 kcal/mol from ΔE(EE) (Figure 5).

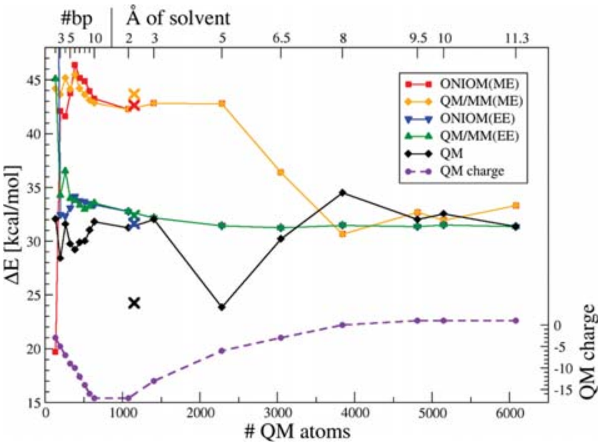


Figure 5. QM size convergence with up to 6000 QM atoms (HF-D3/STO-3G).

On the basis of these (single point) QM size convergence calculations, we suggest that at least five base pairs of the DNA double strand with 5 Å of solvent around these base pairs are necessary within the QM region to reach size convergence within 1 kcal/mol for the present system using electrostatic embedding. This leads to a QM region of about 1000 atoms (Figure 6). Although the exact size of the converged QM region presumably depends on the positions of the counterions, these results are in good agreement with the "big-QM approach" for proteins by Sumner et al.,[8] who suggested the inclusion of all atoms within 4.5 Å of the minimal active site and the movement of junctions at least three residues away. For our DNA system, we find that it is not necessary to include charged groups/ions of the solvent into the QM region in the case of EE. ME and pure QM calculations would benefit from the inclusion of all ions into the QM region; however, such a scheme would not be in line with the typically preferred approach of increasing the QM sphere progressively based on the distance (see also SI-2). The ME scheme can be improved by atomic charges derived from the QM calculation, which recover polarization effects within the QM region.[17,65] Although the QM size convergence behavior can also be improved by this approach, no general scheme to derive QM enhanced charges has been suggested so far that is applicable to QM size convergence (see SI-3). Moreover, none of the tested variations of ME converged faster with the QM size than EE.

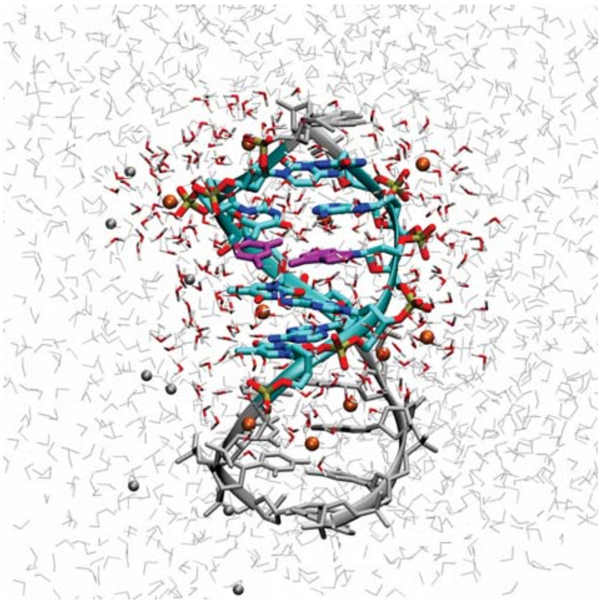


Figure 6. Cutout of the system. The suggested QM region, consisting of five base pairs surrounded by 5 Å of solvent, is shown in color. The protonated base pair (Figure 2) is highlighted in magenta. Atoms treated at the MM level are shown in gray. Protons of the DNA are not shown for clarity. Counterions are represented by balls.

CONCLUSIONS

We showed that the choice of the coupling scheme between the QM and MM parts has no significant influence on the size of the converged QM region; however, the embedding scheme is important. Mechanical embedding, as in the original ONIOM approach, can result in slow convergence with misleading energies, even for large QM sizes, whereas electrostatic embedding leads to fast and reliable convergence of the QM size. No embedding (pure QM) can lead to seemingly fast size convergence that can only be identified as erroneous with very large QM regions. We find that, for the present system, the converged QM size is nearly independent of the chosen method/basis. Therefore, we conclude that QM sizes with no fewer than two adjacent base pairs on each side and at least 5 Å of solvent around these base pairs should ideally be employed for a reliable description of proton transfer processes within a DNA base pair, i.e., a total of 1150 QM atoms.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00727.

Data from Figures 3−5 in tabular form, additional information on the influence of counterions on the QM size convergence of ME and pure QM, additional information on the QM size convergence with QM enhanced charges (PDF)

Coordinates of the systems before proton transfer (PDB)
Coordinates of the systems after proton transfer (PDB)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: christian.ochsenfeld@uni-muenchen.de.

ORCID
Christian Ochsenfeld: 0000-0002-4189-6558

Funding
We acknowledge financial support by the Volkswagen Stiftung within the funding initiative "New Conceptual Approaches to Modeling and Simulation of Complex Systems", and the DFG cluster of excellence EXC 114 "Center for Integrative Protein Science Munich" (CIPSM), and SFB 749.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(2) Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 10020−10031.

(3) Sumowski, C. V.; Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **2009**, *113*, 11734−11741.

(4) Sumowski, C. V.; Schmitt, B. B. T.; Schweizer, S.; Ochsenfeld, C. Quantum-Chemical and Combined Quantum-Chemical/Molecular-Mechanical Studies on the Stabilization of a Twin Arginine Pair in Adenovirus Ad11. *Angew. Chem., Int. Ed.* **2010**, *49*, 9951−9955.

(5) van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010**, *114*, 11303−11314.

(6) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8*, 3793−3803.

(7) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8*, 2260−2271.

(8) Sumner, S.; Söderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4205−4214.

(9) Liao, R.-Z.; Thiel, W. Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comput. Chem.* **2013**, *34*, 2389−2397.

(10) Blank, I. D.; Sadeghian, K.; Ochsenfeld, C. A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination Within the Active Site. *Sci. Rep.* **2015**, *5*, 10369.

(11) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1−28.

(12) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.

(13) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.

(14) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(15) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. Proton Transfer at Metal Sites in Proteins Studied by Quantum Mechanical Free-Energy Perturbations. *J. Chem. Theory Comput.* **2008**, *4*, 985−1001.

(16) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do Quantum Mechanical Energies Calculated for Small Models of Protein-Active Sites Converge? *J. Phys. Chem. A* **2009**, *113*, 11793−11800.

(17) Hu, L.; Söderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761−777.

(18) Du, L.; Gao, J.; Liu, Y.; Zhang, D.; Liu, C. The Reaction Mechanism of Hydroxyethylphosphonate Dioxygenase: A QM/MM Study. *Org. Biomol. Chem.* **2012**, *10*, 1014−1024.

(19) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(20) Löwdin, P.-O. Proton Tunneling in DNA and Its Biological Implications. *Rev. Mod. Phys.* **1963**, *35*, 724−732.

(21) Clementi, E.; Mehl, J.; von Niessen, W. Study of the Electronic Structure of Molecules. XII. Hydrogen Bridges in the Guanine−Cytosine Pair and in the Dimeric Form of Formic Acid. *J. Chem. Phys.* **1971**, *54*, 508.

(22) Scheiner, S.; Kern, C. W. Theoretical Study of Proton Transfers Between Base Pairs of DNA. *Chem. Phys. Lett.* **1978**, *57*, 331−333.

(23) Lipinski, J.; Gorzkowska, E. Double Proton Transfer and Charge-Transfer Transitions in Biological Hydrogen-Bonded Systems. Guanine-Cytosine (G-C) and G-C-Mg2+ Systems. *Chem. Phys. Lett.* **1983**, *94*, 479−482.

(24) Florian, J.; Leszczynski, J. Spontaneous DNA Mutations Induced by Proton Transfer in the Guanine-Cytosine Base Pairs: An Energetic Perspective. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.

(25) Colominas, C.; Luque, F. J.; Orozco, M. Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes. *J. Am. Chem. Soc.* **1996**, *118*, 6811−6821.

(26) Zhanpeisov, N. U.; Leszczynski, J. Specific Solvation Effects on the Structures and Properties of Neutral and One-Electron Oxidized Formamidine-Formamide Complexes. A Theoretical Ab Initio Study. *J. Phys. Chem. A* **1999**, *103*, 8317−8327.

(27) Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. Protonation and Deprotonation Enthalpies of Guanine and Adenine and Implications for the Structure and Energy of Their Complexes with Water: Comparison with Uracil, and Cytosine. *J. Phys. Chem. A* **1999**, *103*, 8853−8860.

(28) Zoete, V.; Meuwly, M. Double Proton Transfer in the Isolated and DNA-Embedded Guanine-Cytosine Base Pair. *J. Chem. Phys.* **2004**, *121*, 4377−4388.

(29) Villani, G. Theoretical Investigation of Hydrogen Transfer Mechanism in the Guanine-Cytosine Base Pair. *Chem. Phys.* **2006**, *324*, 438−446.

(30) Xiao, S.; Wang, L.; Liu, Y.; Lin, X.; Liang, H. Theoretical Investigation of the Proton Transfer Mechanism in Guanine-Cytosine and Adenine-Thymine Base Pairs. *J. Chem. Phys.* **2012**, *137*, 195101.

(31) Noguera, M.; Sodupe, M.; Bertrán, J. Effects of Protonation on Proton-Transfer Processes in Guanine-Cytosine Watson-Crick Base Pairs. *Theor. Chem. Acc.* **2004**, *112*, 318−326.

(32) Noguera, M.; Sodupe, M.; Bertrán, J. Effects of Protonation on Proton Transfer Processes in Watson-Crick Adenine-Thymine Base Pair. *Theor. Chem. Acc.* **2007**, *118*, 113−121.

(33) Chen, H. Y.; Kao, C. L.; Hsu, S. C. N. Proton Transfer in Guanine- Cytosine Radical Anion Embedded in B-Form DNA. *J. Am. Chem. Soc.* **2009**, *131*, 15930−15938.

(34) Matsui, T.; Sato, T.; Shigeta, Y.; Hirao, K. Sequence-Dependent Proton-Transfer Reaction in Stacked GC Pair II: The Origin of Stabilities of Proton-Transfer Products. *Chem. Phys. Lett.* **2009**, *478*, 238−242.

(35) Villani, G. Theoretical Investigation of the Coupling between Hydrogen Atoms Transfer and Stacking Interaction in Guanine-Cytosine Dimers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19242−19252.

(36) Cerón-Carrasco, J. P.; Requena, A.; Zúñiga, J.; Michaux, C.; Perpète, E. A.; Jacquemin, D. Intermolecular Proton Transfer in Microhydrated Guanine-Cytosine Base Pairs: A New Mechanism for Spontaneous Mutation in DNA. *J. Phys. Chem. A* **2009**, *113*, 10549−10556.

(37) Cerón-Carrasco, J. P.; Zúñiga, J.; Requena, A.; Perpète, E. A.; Michaux, C.; Jacquemin, D. Combined Effect of Stacking and Solvation on the Spontaneous Mutation in DNA. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14584−14589.

(38) Matsui, T.; Sato, T.; Shigeta, Y. Sqeuence Dependent Proton-Transfer Reaction in Stacked GC Pair I: The Possibility of Proton-Transfer Reactions. *Int. J. Quantum Chem.* **2009**, *109*, 2168−2177.

(39) Lin, Y.; Wang, H.; Wu, Y.; Gao, S.; Schaefer, H. F., III Proton-Transfer in Hydrogenated Guanine-Cytosine Trimer Neutral Species, Cations, and Anions Embedded in B-Form DNA. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6717−6725.

(40) Jacquemin, D.; Zúñiga, J.; Requena, A.; Céron-Carrasco, J. P. Assessing the Importance of Proton Transfer Reactions in DNA. *Acc. Chem. Res.* **2014**, *47*, 2467−2474.

(41) Cerón-Carrasco, J. P.; Jacquemin, D. DNA Spontaneous Mutation and Its Role in the Evolution of GC-Content: Assessing the Impact of the Genetic Sequence. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7754−7760.

(42) Hays, F. A.; Teegarden, A.; Jones, Z. J. R.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. Correction for Hays et Al., How Sequence Defines Structure: A Crystallographic Map of DNA Structure and Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4486−4486.

(43) Case, D. A.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Hornak, V.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber 11*; University of California: San Francisco, 2010.

(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(45) Wang, J.; Wolf, R.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(46) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(47) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(48) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New Tricks for Modelers from the Crystallography Toolkit: The Particle Mesh Ewald Algorithm and Its Use in Nucleic Acid Simulations. *Structure* **1999**, *7*, 55−60.

(49) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.

(50) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.

(51) Kussmann, J.; Ochsenfeld, C. Pre-Selective Screening for Matrix Elements in Linear-Scaling Exact Exchange Calculations. *J. Chem. Phys.* **2013**, *138*, 134114.

(52) Kussmann, J.; Ochsenfeld, C. Preselective Screening for Linear-Scaling Exact Exchange-Gradient Calculations for Graphics Processing Units and General Strong-Scaling Massively Parallel Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 918−922.

(53) *ChemShell*, a Computational Chemistry Shell. www.chemshell.org.

(54) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113*, 11856−11865.

(55) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell - a Modular Software Package for QM/MM Simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 101−110.

(56) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(57) Bayly, C. C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(58) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(59) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(60) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(61) Retegan, M.; Neese, F.; Pantazis, D. A. Convergence of QM/MM and Cluster Models for the Spectroscopic Properties of the Oxygen-Evolving Complex in Photosystem II. *J. Chem. Theory Comput.* **2013**, *9*, 3832−3842.

(62) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(63) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(64) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functions. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(65) Tao, P.; Fisher, J. F.; Shi, Q.; Vreven, T.; Mobashery, S.; Schlegel, H. B. Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue. *Biochemistry* **2009**, *48*, 9839−9847.

Supporting Information for

Influence of coupling and embedding schemes on the QM size convergence in QM/MM approaches for the example of a proton transfer in DNA.

Sven Roßbach[†,‡] and Christian Ochsenfeld*[,†,‡]

[†] Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU Munich), Butenandtstr. 7, D-81377 Munich, Germany

[‡] Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU Munich), Butenandtstr, 5-13, D-81377 Munich, Germany

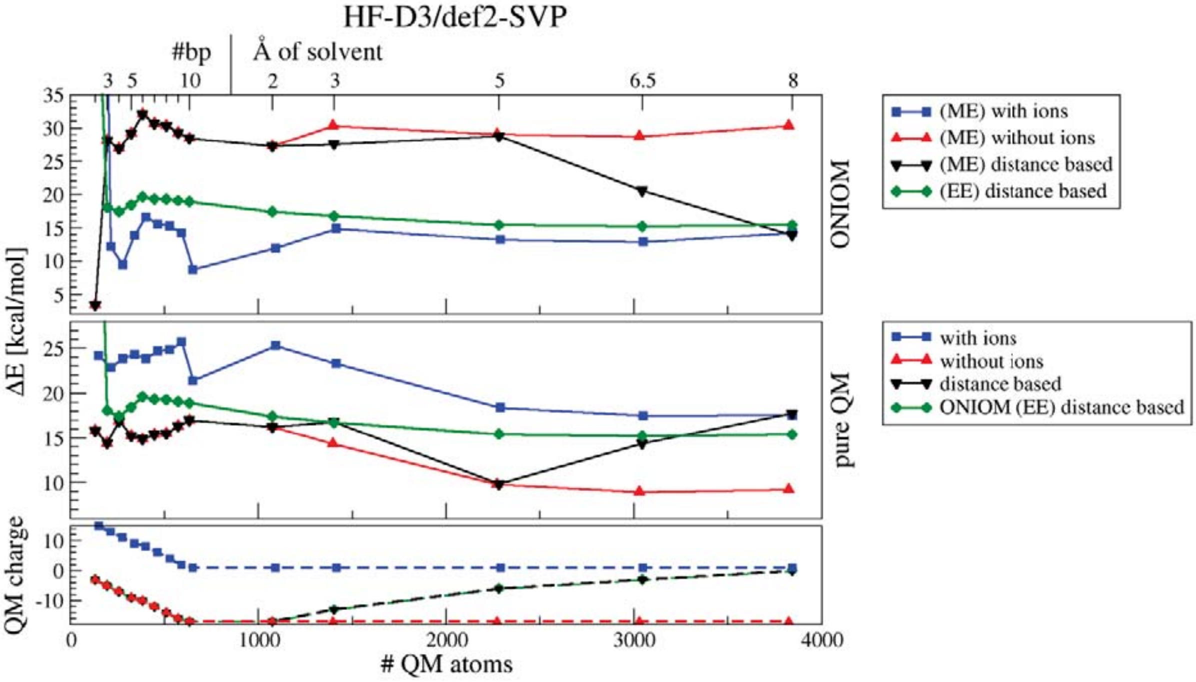
This supporting information contains a table with the values of Fig. 3-5, additional information on the influence of counter ions on the QM size convergence of ME and pure QM and a discussion about mechanical embedding with QM enhanced charges.
The coordinates of the systems before and after the proton transfer are available as pdb files.

1. **SI-Table 1:** Summary of the QM size convergence (Fig 3-5)

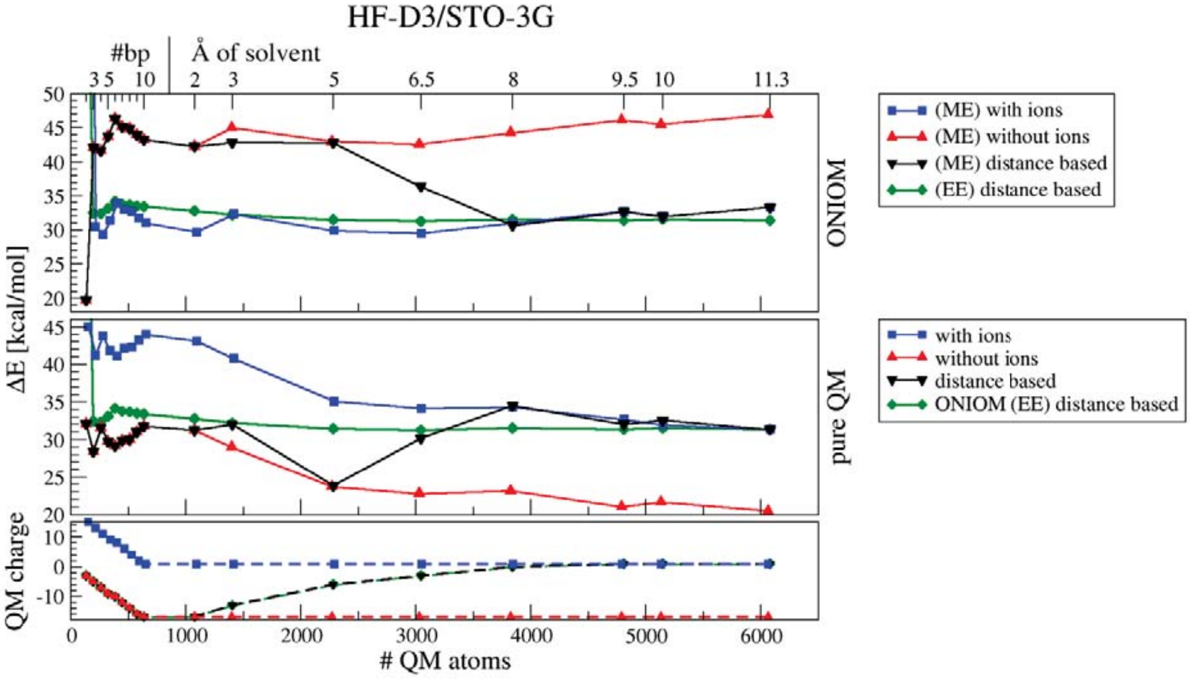
		selection	2bp	3bp	4bp	5bp	6bp	7bp	8bp	9bp	10bp	2Å	3Å	5Å	6.5Å	8Å	9.5Å	10Å	11.3Å
		#atoms	132	196	259	323	383	446	510	573	633	1074	1402	2282	3044	3842	4809	5148	6081
		QM charge	-3	-5	-7	-9	-10	-12	-14	-16	-17	-17	-13	-6	-3	0	1	1	1
B3LYP-D3 / def2-SVP	QM/MM	EE	25,21	16,30	18,02	15,92	15,90	15,51	15,20	15,39	15,62	14,50	13,80	12,45	12,19	12,39			
	ONIOM	EE	13,01	11,19	13,70	12,02	11,72	12,08	12,14	12,88	13,44	13,20	13,91	7,31	11,51	14,64			
	QM/MM	ME	0,64	24,89	23,70	26,02	28,92	27,38	27,02	25,82	24,91	24,21	24,71	26,23	17,71	10,81			
	ONIOM	ME	57,27	14,60	13,97	15,14	16,38	16,03	16,01	15,76	15,57	14,50	13,80	12,45	12,19	12,39			
	QM		13,01	11,19	13,70	12,02	11,72	12,08	12,14	12,88	13,44	13,20	13,91	7,31	11,51	14,64			
M06-2X-D3 / def2-SVP	QM/MM	EE	26,48	17,67	19,36	17,24	17,20	16,83	16,52	16,72	16,95	15,63	14,93	13,60	13,38	13,57			
	ONIOM	EE	58,54	15,96	15,32	16,45	17,68	17,35	17,32	17,08	16,90	15,63	14,93	13,60	13,38	13,57			
	QM/MM	ME	26,32	27,68	28,56	27,67	29,29	27,69	27,09	26,34	25,87	25,37	25,82	27,30	18,86	12,00			
	ONIOM	ME	1,89	26,26	25,07	27,39	30,27	28,78	28,43	27,26	26,36	25,37	25,82	27,30	18,86	12,00			
	QM		14,26	12,56	15,06	13,39	13,07	13,47	13,54	14,31	14,90	14,35	15,02	8,37	12,66	15,84			
HF-D3 / STO3G	QM/MM	EE	45,11	34,27	36,52	33,98	33,78	33,35	32,99	33,25	33,56	32,77	32,18	31,45	31,25	31,50	31,37	31,52	31,36
	ONIOM	EE	77,05	32,46	32,36	33,09	34,15	33,76	33,69	33,50	33,40	32,77	32,18	31,45	31,25	31,50	31,37	31,52	31,36
	QM/MM	ME	44,21	43,63	45,19	44,12	45,51	44,19	43,65	43,14	42,86	42,29	42,85	42,81	36,40	30,67	32,70	31,98	33,33
	ONIOM	ME	19,70	42,13	41,62	43,76	46,41	45,19	44,90	43,97	43,27	42,29	42,85	42,81	36,40	30,67	32,69	31,98	33,33
	QM		32,07	28,43	31,61	29,76	29,21	29,89	30,02	31,03	31,80	31,27	32,04	23,88	30,21	34,51	32,05	32,57	31,36
HF-D3 / def2-SVP	QM/MM	EE	28,47	19,76	21,45	19,21	19,14	18,79	18,47	18,69	18,95	17,41	16,71	15,41	15,21	15,38			
	ONIOM	EE	60,55	18,07	17,41	18,43	19,62	19,32	19,28	19,06	18,90	17,41	16,71	15,41	15,21	15,38			
	QM/MM	ME	27,83	29,51	30,37	29,51	31,06	29,59	29,03	28,36	27,94	27,25	27,59	28,77	20,61	13,91			
	ONIOM	ME	3,44	28,13	26,91	29,25	32,07	30,70	30,39	29,30	28,45	27,25	27,59	28,77	20,61	13,91			
	QM		15,81	14,43	16,91	15,25	14,87	15,40	15,51	16,35	16,99	16,24	16,79	9,85	14,41	17,75			

2. Influence of counter ions

It can be argued that the slow convergence of mechanical embedding and pure QM compared to electrostatic embedding is caused by the counter ions. This can be examined by including and excluding all counter ions during the size convergence. SI-Figure 1 and 2 show the results of these calculations for HF-D3 with the def2-SVP basis set and STO-3G, respectively. The case without ions in the QM region converges to a different energy but, the size convergence shows the same trend when increasing the amount of solvent to the QM region like the case with all ions included from the beginning. When counter ions and water molecules are progressively included based on their distance to the DNA, the energy difference first follows the case without ions and then meets the results with all ions included. The inclusion of all counter ions into the QM region smoothens the convergence of mechanical embedding and pure QM. Nevertheless, the QM size convergence is still significantly slower in both cases compared to electrostatic embedding.



SI-Figure 1: Influence of counter ions on the QM size convergence of the ONIOM (mechanical embedding) method and pure QM on HF-D3/def2-SVP level of theory. First, base pairs are added to the QM region until all 10 base pairs are included, then water and counter ions are added based on the distance to the DNA (black). The calculations have been repeated with all counter ions included in the QM region from the beginning (blue) and without any counter ions in the QM region (red). In green, the results of ONIOM (electrostatic embedding) are shown as reference.

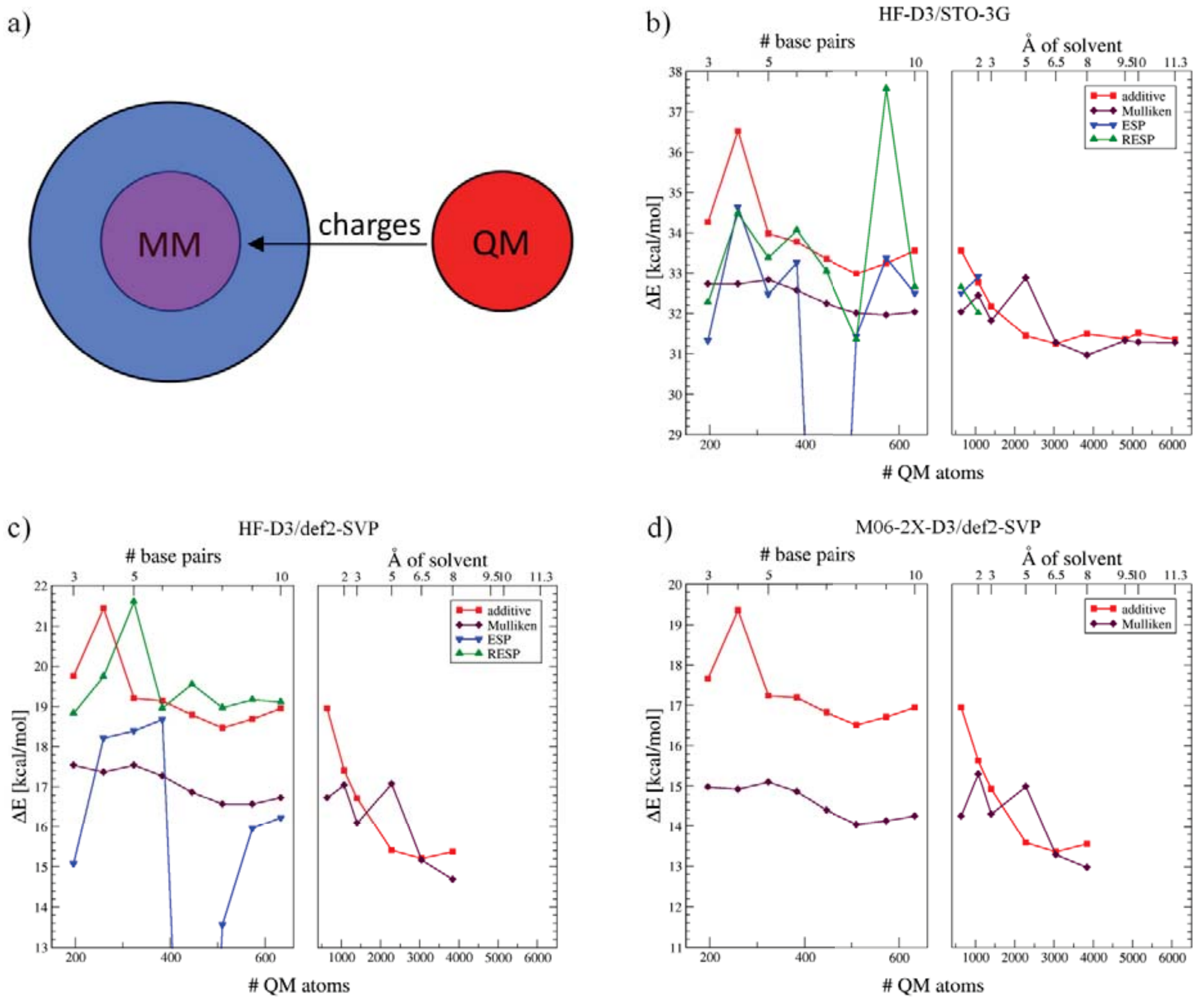


SI-Figure 2: Influence of counter ions on the QM size convergence of the ONIOM (mechanical embedding) method and pure QM on HF-D3/STO-3G level of theory with up to 6000 QM atoms. First, base pairs are added to the QM region until all 10 base pairs are included, then water and counter ions are added based on the distance to the DNA (black). The calculations have been repeated with all counter ions included in the QM region from the beginning (blue) and without any counter ions in the QM region (red). In green, the results of ONIOM (electrostatic embedding) are shown as reference.

3. Mechanical embedding with QM enhanced charges

The main disadvantage of ME is the lack of mutual polarization between QM and MM region. Apart from treating the electrostatic interaction on the QM level (EE), its description can also be improved in the ME scheme. If the QM calculation is performed first, atomic charges can be derived and the corresponding MM charges updated (SI-Figure 3a), recovering the polarization within the QM region.[1,2] Mulliken charges can be obtained basically for free and they perform surprisingly well for the chosen system, method and basis set combinations (SI-Figure 3b-d). Nevertheless, Mulliken charges are known to heavily depend on the method and basis set[3–5] and cannot even be converged with the size of the basis set.

Charges that are fitted to the electrostatic potential of the QM density (ESP charges), which is evaluated on a grid, strongly depend on the charge fitting procedure. Buried atoms are a particular problem because their charge is shielded by others, reducing their influence on the resulting electrostatic potential. Unfortunately, the number of buried atoms increases rapidly when the QM region is increased radially. Restrained electrostatic potential (RESP) charges avoid this problem but are also dependent on the charge fitting procedure. With a minimal basis set, both, ESP and RESP charges, lead to large energy fluctuations with increasing QM size (SI-Figure 3b). With a larger basis set, we find RESP charges to perform reasonably well (SI-Figure 3c). However, the iterative fitting procedure soon dominates the calculation. Nevertheless, other charge derivation schemes based on fragmentation or deformation density may be better suited.



SI-Figure 3: a) Visualization of the QM enhanced charges in a mechanical embedding scheme. b-d) QM size convergence (ONIOM(ME)) with different method/basis set combinations and different charge derivation schemes. The results of the additive QM/MM scheme with electrostatic embedding are shown in red for comparison. The results with Mulliken charges and charges fitted to the electrostatic potential (ESP) and the restraint electrostatic potential (RESP) are shown in purple, blue and green, respectively.

References

(1) Tao, P.; Fisher, J. F.; Shi, Q.; Vreven, T.; Mobashery, S.; Schlegel, H. B. Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue. *Biochemistry* **2009**, *48*, 9839–9847.

(2) Hu, L.; Söderhjelm, P.; Ryde, U. On the Convergence of QM/MM Energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

(3) Guerra, C. F.; Handgraaf, J.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and

VDD Methods for Charge Analysis. *J. Comput. Chem.* **2004**, *25*, 189–210.

(4) Martin, F.; Zipse, H. Charge Distribution in the Water Molecule -A Comparison of Methods. *J. Comput. Chem.* **2005**, *26*, 97–105.

(5) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Assessment of Atomic Charge Models for Gas-Phase Computations on Polypeptides. *J. Chem. Theory Comput.* **2012**, *8*, 661–676.

Paper 3: “Effect of DNA Environment on Electronically Excited States of Methylene Blue in a QM/QM/MM Scheme”, J. J. Nogueira, S. Roßbach, C. Ochsenfeld, L. González, *to be submitted*.

Effect of DNA Environment on Electronically Excited States of Methylene Blue in a QM/QM/MM Scheme

Juan J. Nogueira[†,1], *Sven Roßbach*[‡,1], *Christian Ochsenfeld**[,‡], *Leticia González**[,†]

[†]Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna

Währinger Str. 17 A-1090 Wien, Austria.

[‡]Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstrasse 7, D-81377 Munich, Germany.

[1]These authors contribute equally to the manuscript.

Corresponding Author

* christian.ochsenfeld@cup.uni-muenchen.de, *leticia.gonzalez@univie.ac.at

KEYWORDS: ONIOM QM/QM/MM, Electronically Excited States, Biological Environment, Methylene Blue, DNA.

ABSTRACT. Interactions between chromophores and biological environments can induce important modifications on the electronic properties of the chromophores. A three-layered ONIOM scheme with electrostatic embedding is implemented to investigate the influence of an additional layer, with respect to a standard QM/MM description, to treat a part of the environment at quantum mechanical level in excited-state calculations. The approach is employed to compute the electronic excitations of the photosensitizer methylene blue interacting with a solvated DNA strand. It is shown that the quantum mechanical description of several nucleobases induces energy shifts in the excited states of methylene blue, compared to a QM/MM scheme with a fully classical description of the solvated double strand. The energy shifts present a polarizing electrostatic component, caused by a charge redistribution of the environment, and an electronic coupling component, originated by the mixing between the electronic bright state of methylene blue and a charge-transfer state between methylene blue and guanine. While the electrostatic effect can also be taken into account by additive schemes, the description of electronic coupling is only possible by employing a subtractive scheme as the one used here.

1. INTRODUCTION

Hybrid quantum mechanics/molecular mechanics (QM/MM) approaches[1-3] have been established as valuable tools for modeling physical and chemical processes in biological systems occurring in both electronic ground and excited states. Since excited electrons are less tightly held by the nucleus than electrons in the electronic ground state, the effect of the biological environment on electronic properties is generally more important in excited-state processes. QM/MM calculations have been widely employed in the literature to investigate the photophysics and photochemistry of chromophores embedded in solution[4-10] and in biological environment.[9,11-19] Depending on the process under investigation, the environment may play an active or passive role on the photoinduced event. In the first case, the fragment of the environment involved in the process needs to be included in the QM region at the same level of theory as the chromophore. This recipe was followed, for example, when studying electron transfer between nucleobases in DNA,[15] photoinduced thymine dimerization,[12] and electron-driven proton transfer from 2-aminooxazole[6] and 5,6-dihydroxyindole[20] to solvent molecules. In the case of a passive participation, the environment is, in general, completely described by a force field in the MM region, which is able to induce a shift in the potential energy surfaces of the QM region, tuning its electronic properties. This description was successfully employed to show that, e.g., the excited-state lifetimes of adenine are one order of magnitude higher in DNA strands than in water,[17] intersystem crossing efficiency in methylene blue is affected by the presence of DNA[11] or lipid environments,[21] and the charge-transfer character of the lowest-energy UV absorption band of nitrobenzene decreases when going from the gas-phase to aqueous environment.[4]

The most common QM/MM approach employed in the literature involves electrostatic embedding,[2] in which the point charges of the MM region are included in the QM Hamiltonian as one-electron operators. Thus, the wavefunction of the QM region is polarized by the MM point charges, but the electronic structure of the environment is not affected by the presence of the QM region due to the use of fixed-charge force fields. This approach provides accurate results in many cases, although it can fail when the excitations of the chromophore are strongly coupled to the environment by interactions that are not purely electrostatic. This is the case, for example, for the green fluorescent protein, whose absorption spectrum calculated by a QM/MM electrostatic scheme is blue-shifted with respect to the experimental spectrum.[14] This energy overestimation was alleviated by the inclusion of induced dipole moments in the MM region, allowing the environment to respond to the electronic alterations of the chromophore after excitation due to the polarizable QM/MM embedding.[22] A different strategy to improve the description of the environment is to use a QM method to treat the relevant part of the environment that is strongly interacting with the chromophore. This can be achieved by a partitioning into three layers: The chromophore is in the first layer, described by a high-level QM method, part of the environment is in the second layer, described by a medium-level QM method, and the rest of the environment is in the third layer, treated by a low-level method, e.g. a force field. In this spirit, the excited states of two rhodopsin receptors were theoretically investigated by a QM1/QM2/MM approach employing an iterative fitting procedure to compute the mutual polarization between the QM layers.[23] In this formulation, electrostatic interactions are more accurate than those calculated by a standard QM/MM electrostatic approach, but non-electrostatic effects (e.g. dispersion and electronic exchange) are still neglected. The more accurate electronic distribution of the environment, due to its quantum mechanical description, resulted in a red-shift of the rhodopsin excited states of only 0.1 eV with respect to a full classical description of the environment, which was explained by the limited size of the QM2 region.[23] One of the benefits of the three-layered QM1/QM2/MM approach over the QM/MM polarizable embedding is that the use of polarizable force fields, whose availability is still limited, is not required.

In the present work, we employ a three-layered QM1/QM2/MM approach using the ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) scheme with electrostatic embedding (ONIOM3-EE).[24] The subtractive ONIOM scheme has one clear advantage over additive schemes. While additive approaches can only describe excitations located at the chromophore (local excitations), the ONIOM approach can, in principle, describe excited states that are delocalized over both the chromophore and the environment, for example, intermolecular charge-transfer and Frenkel exciton electronic states. These delocalized excited states are described in the ONIOM scheme by the medium-level method employed in the QM2 layer. The influence of the increased QM2 region is therefore extrapolated at a lower QM level. We apply here the developed ONIOM3-EE methodology to the calculation of the low-lying electronically excited states of methylene blue when it is interacting with a solvated DNA double strand. Methylene blue is a photosensitizer that has shown promising results in the treatment of tumors[25-27] and microbial infections.[28,29] It is known that methylene blue is able to interact with DNA strands and its binding modes have been well characterized by spectroscopic techniques[30-32] and theoretical simulations.[33,34] Recently, it has been shown that the environment has a significant influence on the intersystem crossing mechanism of methylene blue, which is enhanced when methylene blue is intercalated into DNA.[11] Thus, methylene blue/DNA complexes are good candidates to test our three-layered QM1/QM2/MM approach and investigate the effect of the additional QM2 layer on the excited-state electronic structure of the chromophore. In addition, the measured absorption spectrum[35,36] and other spectroscopic properties[36] of methylene blue integrated in DNA are available, allowing comparison with experimental data. The remainder of the article is organized as follows: the ONIOM3-EE approach for excited-state calculations and computational details are described in Section 2. The calculated excited-state energies of methylene blue integrated in a DNA strand are presented and discussed in Section 3. Specifically, the effect of including part of the DNA strand at quantum mechanical level during the excited-state calculations and geometry optimizations are discussed in Sections 3.1 and 3.2, respectively. Finally, concluding remarks are drawn in Section 4.

2. THEORY AND MODEL

2.1 QM1/QM2/MM ONIOM Electrostatic Embedding Approach. In the ONIOM3 approach,[37] the system is split into three layers, which are described at different levels of theory, as is illustrated in Figure 1. The most relevant part of the system, in our case the chromophore, is termed model system and is described by the highest level of theory (QM1) in the first layer. The second layer comprises the chromophore and part of the environment, the intermediate model system, and is treated with a medium-level method (QM2). The real system, which includes everything, is modeled by MM using a force field. The ONIOM ground-state energy is calculated by a subtractive scheme as:[24]

$$E^{ONIOM} = E^{MM}_{real} + E^{QM2,v}_{int-model} - E^{MM,v}_{int-model} + E^{QM1,v}_{model} - E^{QM2,v}_{model} \qquad (1)$$

where the negative terms avoid double counting of the int-model and model layers. The superscript v indicates the use of electrostatic embedding by including the point charges of the fragment of the system that is not being considered in the QM calculation. For example, in the term $E^{QM1,v}_{model}$ the energy of the model subsystem is calculated at QM1 level including the point charges of the second and third layer in the QM1 Hamiltonian. In a similar way, excitation energies ΔE^{ONIOM} can be calculated in the spirit of the ONIOM approach as:[38]

$$\Delta E^{ONIOM} = \Delta E^{MM}_{real} + \Delta E^{QM2,v}_{int-model} - \Delta E^{MM,v}_{int-model} + \Delta E^{QM1,v}_{model} - \Delta E^{QM2,v}_{model} \tag{2}$$

Since the force field is not able to describe electronically excited states, the first and third term on the right side of equation (2) vanish ($\Delta E^{MM}_{real} = 0$ and $\Delta E^{MM,v}_{int-model} = 0$). Thus, equation (2) can be rewritten as:

$$\Delta E^{ONIOM} = \Delta E^{QM2,v}_{int-model} + \Delta E^{QM1,v}_{model} - \Delta E^{QM2,v}_{model} \tag{3}$$

Although the force field cannot calculate excitation energies, its use in the ONIOM calculation is still necessary in order to take into account the point charges of the environment in the calculation of the three energy terms of the right side of equation (3). We have implemented this three-layered ONIOM scheme to calculate excited-state energies in the ChemShell interface.[39-41]

In previous excited-state calculations for rhodopsin, using a three-layered ONIOM approach, it was shown to be necessary to include the point charges from the real system not only into the calculation of the int-model, but also into the calculation of the model system.[24] Later, it was found to be beneficial to embed the model layer in the field created by the point charges of both other layers.[38] Fixed charges of the force field were used to represent the int-model layer. Wanko et al.[23] investigated the effect of mutual polarization of the int-model and model systems by iteratively fitted RESP charges. The effect on the excitation energy was only moderate, which was explained by the limited size of the outer QM region. Very recently, Biancardi et al.[42] investigated the influence of QM-derived charges on excitation energies and oscillator strengths in a two layer QM/QM-EE scheme. Although the main focus lay on the influence of link atoms, it was concluded that the polarization response of the embedding point charges to electronic excitation should improve the results.

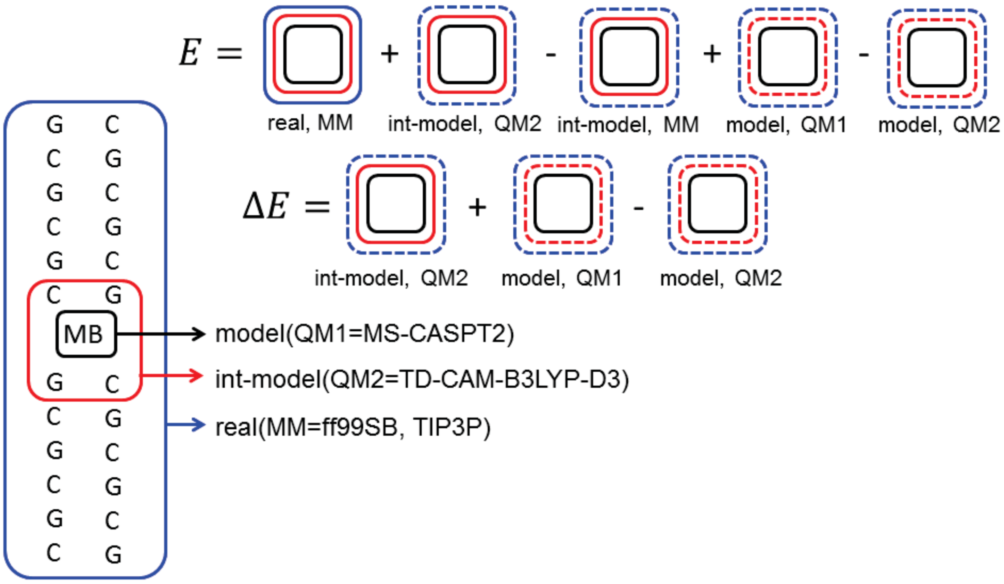


Figure 1. Schematic representation of the three-layered ONIOM scheme developed in this work to calculated ground-state energies E and excited-state energies ΔE. The model, intermediate model and real systems are represented by black, red and blue squares, respectively. Dashed-line squares represent layers whose point charges are included in the electrostatic embedding scheme during the energy calculation of the layers represented by solid-line squares. The left-side scheme represents the system investigated here, namely methylene blue integrated into a (dG-dC)$_6$ double strand.

Here, in our ONIOM3-EE scheme, the description of the int-model layer is improved by Mulliken charges computed by the QM2 method. In this way, the int-model system is polarized by the environment and the model system is polarized by the static point charges from the environment and the QM-derived charges of the int-model system. The link-atom approach with the charge-shift scheme[39] is employed to treat covalent bonds across layers. In the link-atom approximation, the point charge of the low-level atom, connected to a high-level atom, can lead to overpolarization at the high-level border. This problem is often avoided by setting to zero the charges of the MM atoms close to the border.[38] However, such a procedure strongly modifies the charge distribution of the modified atoms in an unrealistic manner. A better way to overcome the overpolarization problem is to employ the charge-shift scheme,[39] which moves the point charge of the bordering low-level atom to the connected low-level atoms. The resulting dipole moment is corrected by additional dipole moments at the position of these atoms to preserve the real dipole moment of the system.

2.2 Computational Details. The first step to test our ONIOM approach is to obtain reliable structures for methylene blue interacting with a DNA strand. Previous classical molecular dynamics (MD) simulations have shown that methylene blue presents two energetically favorable binding modes when it interacts with a $(dGdC)_6$ double strand, namely intercalative and minor-groove binding modes.[33] We have selected the last snapshot for each binding mode from these previous simulations[33] and classically minimized their geometries using the ff99SB[43,44] and TIP3P[45] force fields. The minimizations were evolved employing the steepest descent method for 10000 steps and conjugate gradient method for 10000 additional steps as implemented in the Amber14 package.[46] The Coulomb and van der Waals' interactions were truncated at 10 Å and the particle mesh Ewald (PME) method[47] was employed to compute the Coulomb interactions. The classically minimized geometries were further refined by an electrostatic embedded QM/MM scheme as implemented in ChemShell.[39-41] Two optimizations with different sizes of the QM region were performed. In the first optimization, only methylene blue was described quantum mechanically at density functional theory (DFT) level using the

B3LYP functional[48] with dispersion correction[49] and the 6-31G(d) basis set[50] as implemented in the Q-Chem 4.2 quantum chemistry program.[51] The linear-scaling screening for computing the exchange part of the Fock matrix was employed.[52] The $(dGdC)_6$ strand and the water molecules were described by a force field[43-45] through the DL_POLY implementation of ChemShell.[39-41] The minor groove (min) and intercalated (int) geometries computed in this way will be named min_{small} and int_{small}, respectively, indicating these geometries have been optimized with the smaller QM region. In the second geometry optimization, the QM region comprises methylene blue and the 4 flanking nucleobases (92 atoms) for the intercalative geometry and methylene blue and the 8 closest nucleobases (146 atoms) for the minor-groove geometry. The geometries obtained by this procedure will be named min_{large} and int_{large} along the manuscript to indicate that a larger QM region was used. The size of the QM region during the geometry optimizations were chosen based on the convergence of excited-state energies when a different number of nucleobases are included in the second layer of ONION3-EE calculations (see below). For both geometry optimizations, the atoms separated from any atom of methylene blue by a distance of 10 Å or less are allowed to move. This set up results in 1388 and 1311 active atoms for the minor-groove and intercalative structures, respectively. The two QM/MM setups employed during the geometry optimizations are represented in Figure 2.

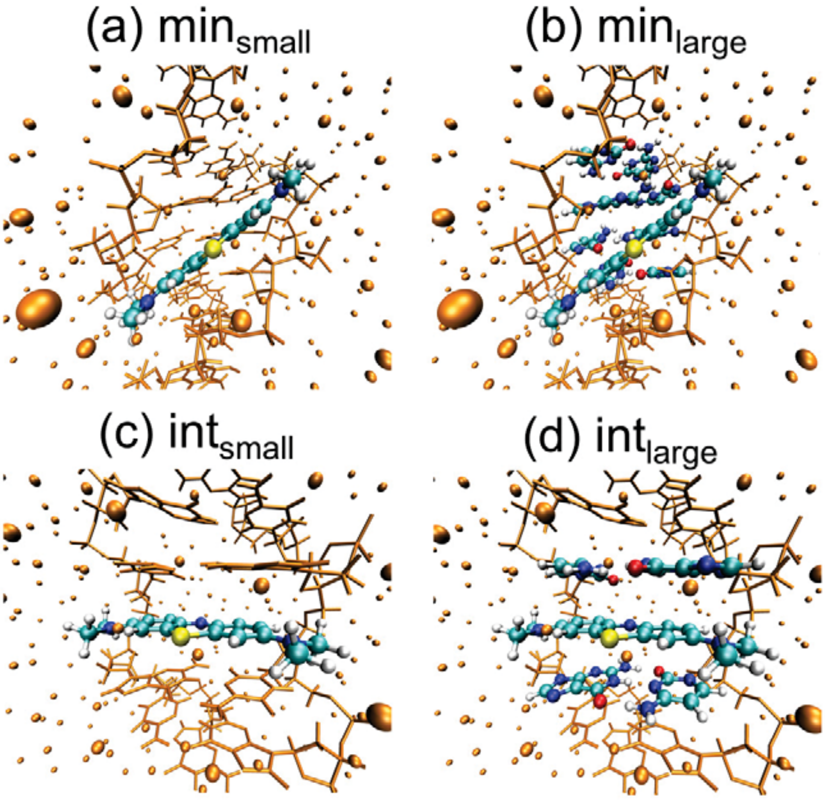


Figure 2. Active atoms of the optimized geometries used in the excited-state energy calculations for the minor-groove geometry including (a) methylene blue and (b) methylene blue and the 8 closest nucleobases in the QM region, the intercalative geometry including (c) methylene blue and (d) methylene blue and the 4 flanking nucleobases in the QM region. C, N, S, O and H atoms of the QM region are cyan, blue, yellow, red and white, respectively. The active residues of the MM region are represented by orange sticks (DNA) and bubbles (water). The frozen MM atoms are not displayed for simplicity.

The four optimized geometries displayed in Figure 2 were employed for electronically excited-state calculations. Again, two types of calculations were performed with different number of atoms described quantum mechanically. In the first set of excited-state calculations, a standard QM/MM electrostatic scheme was employed, where only methylene blue was included in the

QM region and described by multistate complete active space second-order perturbation theory (MS-CASPT2)[53] using the ANO-RCC-VDZP basis set.[54] The active space of 14 electrons and 12 orbitals described in Ref. 11 was employed and 8 roots with equal weights were included in the multistate calculation. To minimize the effect of intruder states, the level-shift approach was applied with a real-valued shift of 0.3 a.u.. The IPEA shift was set to zero, as previously recommended for organic chromophores.[55] These calculations were conducted with a development version of the ChemShell[39-41] interface, which allows the use of MOLCAS 8.0[56] to compute the excited-state energies of the QM region. The DNA strand and water molecules were described by a force field.[43-45] The excited-state energies obtained by this two-layered QM/MM approach, where only methylene blue is described quantum mechanically, will be termed hereinafter $\Delta E^{QM/MM}$. In the second set of computations, the ONIOM3-EE scheme, explained above and implemented in ChemShell, was employed. Specifically, methylene blue was described at MS(8)-CASPT2(14,12)/ANO-RCC-VDZP level of theory[53-55] in the QM1 layer with the same computational details seen above. The 8 and 4 closest nucleobases from the photosensitizer for the minor-grove and intercalative structures, respectively, were included in the QM2 layer and calculated by time-dependent DFT (TD-DFT) using the CAM-B3LYP functional[57] and 6-31G(d) basis set.[50] The rest of the system was described in the MM layer by a force field.[43-45] The excitation energies provided by the three-layered ONIOM approach are termed ΔE^{ONIOM} along the manuscript.

The combination between the different recipes employed for geometry optimizations and excited-state calculations allow us to investigate individually the effect of describing part of the environment quantum mechanically during the excited-state calculation (Section 3.1) and during the geometry optimization process (Section 3.2). Such an analysis will be performed for both intercalative and minor-groove binding modes to study whether environmental effects on the excitation energies of methylene blue depend on the binding site. We will focus our discussion on the excitation energies of $S_1(\pi_1\pi_1^*)$, $T_2(\pi_2\pi_1^*)$ and $T_3(n\pi_1^*)$, which are the electronic states involved in intersystem crossing of methylene blue.[11] The molecular orbitals that participate in the electronic transitions of these states are plotted in Figure 3.

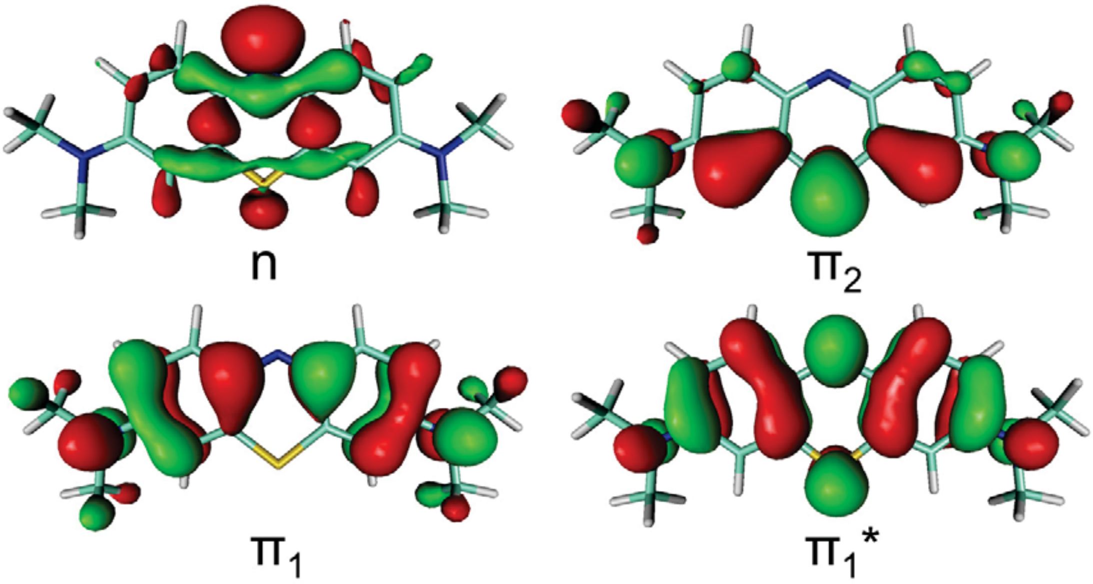


Figure 3. Molecular orbitals involved in the main electronic transitions of the electronically excited states $S_1(\pi_1\pi_1^*)$, $T_2(\pi_2\pi_1^*)$ and $T_3(n\pi_1^*)$.

3. RESULTS AND DISCUSSION

3.1 Excited-State Calculations with Quantum Mechanical Environment. In this section the effect of describing part of the DNA strand at quantum mechanical level during the excited-state calculation will be discussed. To do that, excited-state calculations conducted by the ONIOM3-EE approach explained in Section 2.1 and by the standard two-layered QM/MM approach were performed. Both types of calculations were carried out at the $\text{min}_{\text{small}}$ and $\text{int}_{\text{small}}$ optimized geometries. The first step of our analyses is to investigate the excited-state energy convergence for the ONIOM3-EE calculations by including different number of nucleobases in the second layer. Specifically, excited-state calculations including 1 to 12 nucleobases in the QM2 region were carried out. The variation of the difference between the ONIOM3-EE excitation energy (ΔE^{ONIOM}) and the QM/MM excitation energy ($\Delta E^{QM/MM}$) for the bright $S_1(\pi_1\pi_1{}^*)$ state with the size of the QM2 layer is plotted in Figure 4a,b. For the minor-groove geometry, the ONIOM3-EE excitation energy decreases by 0.07 eV for the largest QM2 region with respect to the QM/MM energy, and the calculation is already converged when 8 nucleobases are included in the second layer. For the intercalative geometry, the excited-state energy change is larger (-0.26 eV), and the calculation is converged when 4 nucleobases are described in the QM2 region. The calculated energy shift, when going from the QM/MM scheme to the ONIOM3-EE scheme, has two different contributions: polarizing electrostatic contribution and electronic coupling contribution. The physical origin of both contributions and their importance for the different binding modes will be discussed in detail below. For now, we only analyze the convergence of the two contributions in Figure 4a,b. As can be seen, the polarizing electrostatic contribution is more important than the electronic-coupling one for the minor-groove geometry, while the opposite stands for the intercalative geometry. Both energy-shift contributions are converged after including 4 and 8 nucleobases in the QM2 layer for the intercalative and minor-groove structures, respectively. Thus, the two different energy contributions show similar convergence behavior as the excitation energies. Once we have seen that the ONIOM3-EE calculations are converged, we analyze in more detail the energy differences between the two-layered QM/MM calculations and the ONIOM3-EE calculations with 8 nucleobases in QM2 for the minor-groove geometry and 4 nucleobases in QM2 for the intercalative geometry.

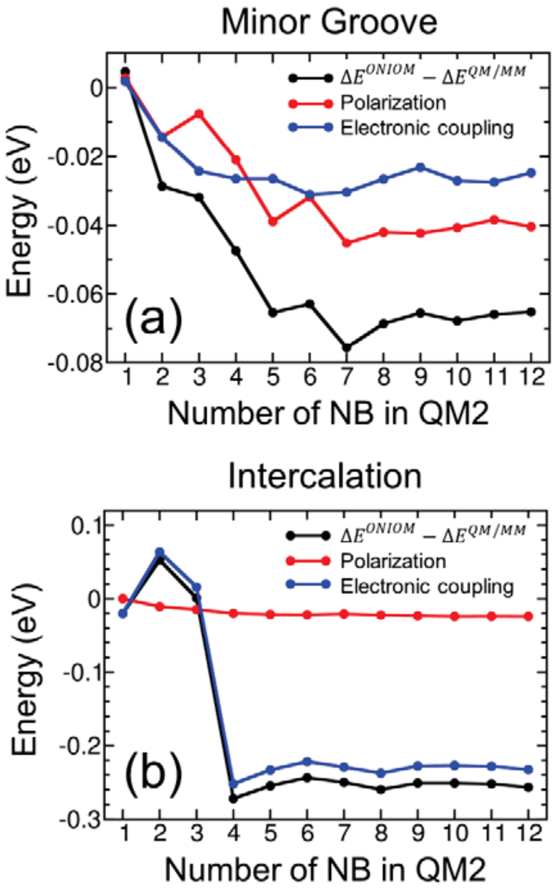


Figure 4. Variation of the difference between ONIOM excited-state energy and QM/MM excited-state energy for $S_1(\pi_1\pi_1*)$ ($\Delta E^{ONIOM} - \Delta E^{QM/MM}$) (black line), polarizing electrostatic contribution (red line) and electronic-coupling contribution (blue line) with the number of nucleobases (NB) in the QM2 layer for (a) the $\mathrm{min}_{\mathrm{small}}$ geometry and (b) the $\mathrm{int}_{\mathrm{small}}$ geometry.

The computed excitation energies ($\Delta E^{ONIOM}$and $\Delta E^{QM/MM}$) and the three ONIOM contributions (see equation 3) $\Delta E^{QM2,v}_{int-model}$, $\Delta E^{QM1,v}_{model}$ and $\Delta E^{QM2,v}_{model}$ are given in Table 1 for the minor-groove and intercalative structures. In the system investigated here, the model system is formed by methylene blue (MB) and the int-model system by methylene blue and 4 (8) nucleobases (MB+DNA) for the intercalative (minor-groove) structure. The two levels of theory QM1 and QM2 are CASPT2 and CAM-B3LYP, respectively, as can be seen in Figure 1. For the sake of clarity, the three ONIOM contributions $\Delta E_{int-model}^{QM2,v}$, $\Delta E_{model}^{QM1,v}$ and $\Delta E_{model}^{QM2,v}$ are renamed and appear in Table 1 as ΔE_{MB+DNA}^{CAM}, ΔE_{MB}^{CASPT2} and ΔE_{MB}^{CAM}. The excited-state energy differences ($\Delta E^{ONIOM} - \Delta E^{QM/MM}$) between both approximations listed in Table 1 show that the description of part of the DNA strand by TD-DFT introduce variations in the excitation energies smaller than 0.1 eV, except for the $S_1(\pi_1\pi_1{}^*)$ state of the intercalative geometry, whose energy shift is -0.26 eV. In the following, the origin of these energy differences, when comparing both methods, will be discussed paying special attention to the large red shift of the $S_1(\pi_1\pi_1{}^*)$ state.

Table 1. Excited-state energies (in eV) of methylene blue located into the minor-groove and intercalative pockets of a $(dGdC)_6$ double strand. Energies obtained by the two-layered QM/MM ($\Delta E^{QM/MM}$) and three-layered ONIOM (ΔE^{ONIOM}) schemes together with the different ONIOM contributions ΔE_{MB}^{CASPT2}, ΔE_{MB}^{CAM}, and ΔE_{MB+DNA}^{CAM} (see equation 3 and text) and the difference between both approximations ($\Delta E^{ONIOM-QM/MM}$) are listed.

$\text{min}_{\text{small}}$ geometry						
	$\Delta E^{QM/MM}$	ΔE^{ONIOM}	ΔE_{MB}^{CASPT2}	ΔE_{MB}^{CAM}	ΔE_{MB+DNA}^{CAM}	$\Delta E^{ONIOM} - \Delta E^{QM/MM}$
$S_1(\pi_1\pi_1{}^*)$	1.99	1.92	1.95	2.59	2.56	-0.07
$T_2(\pi_2\pi_1{}^*)$	2.04	1.97	1.97	2.04	2.04	-0.07
$T_3(n\pi_1{}^*)$	3.05	2.96	3.05	3.12	3.03	-0.09
$\text{int}_{\text{small}}$ geometry						
	$\Delta E^{QM/MM}$	ΔE^{ONIOM}	ΔE_{MB}^{CASPT2}	ΔE_{MB}^{CAM}	ΔE_{MB+DNA}^{CAM}	$\Delta E^{ONIOM} - \Delta E^{QM/MM}$
$S_1(\pi_1\pi_1{}^*)$	1.98	1.72	1.97	2.65	2.40	-0.26
$T_2(\pi_2\pi_1{}^*)$	1.91	1.84	1.88	2.02	1.98	-0.07
$T_3(n\pi_1{}^*)$	2.72	2.74	2.69	2.72	2.77	0.02

When the neighboring nucleobases of methylene blue are described at quantum mechanical level, two different effects, namely polarizing electrostatic effect and the electronic coupling effect, induce alterations in the excitation energies. The former is a consequence of the variation of the charge distribution surrounding the photosensitizer. In the two-layered QM/MM approach, the charges of the atoms of the environment that are included in the Hamiltonian during the CASPT2 calculation are given by the force field. Since a fixed-charge force field is employed here, the charges of the environment are not allowed to change. However, in the ONIOM3-EE approach, part of DNA is described quantum mechanically at TD-DFT level, and thus the charges of this fragment of DNA are computed quantum mechanically taken into account the presence of methylene blue and the rest of the environment. As a consequence, the charges of the environment that are included in the Hamiltonian during the CASPT2 calculation can significantly differ from those of the force field. This electrostatic contribution to the excited-state energies can be computed as the difference between the CASPT2 energies of the ONIOM calculations and the CASPT2 energies of the QM/MM calculations (Table 2). For example, for the $S_1(\pi_1\pi_1^*)$ state of the minor-groove geometry, the electrostatic contribution to the shift of the excitation energy is 1.95 - 1.99 = - 0.04 eV. The electrostatic effect could induce either red shift or blue shift in the excitation energies depending on the degree of stabilization or destabilization of the different electronic states due to the change in the charge distribution of the environment. As seen in Table 2, the excitation energies of the electronic states $S_1(\pi_1\pi_1^*)$, $T_2(\pi_2\pi_1^*)$, and $T_3(n\pi_1^*)$ for the $\text{int}_{\text{small}}$ geometry and the states $S_1(\pi_1\pi_1^*)$ and $T_2(\pi_2\pi_1^*)$ for the $\text{min}_{\text{small}}$ geometry are red shifted, indicating that the excited states are more stabilized than the ground state. The energy of the $T_3(n\pi_1^*)$ state for $\text{min}_{\text{small}}$ does not suffer any modification indicating that both $T_3(n\pi_1^*)$ and the ground state are energetically stabilized in the same extent.

Table 2. Polarizing contribution (in eV), computed as $\Delta E_{MB}^{CASPT2} - \Delta E^{QM/MM}$, and electronic coupling contribution (in eV), computed as $\Delta E_{MB+DNA}^{CAM} - \Delta E_{MB}^{CAM}$, to the shift in the excitation energies of the minor-groove ($\text{min}_{\text{small}}$) and intercalative ($\text{int}_{\text{small}}$) geometries.

	Polarizing Effect		Electronic Coupling Effect	
	$\text{min}_{\text{small}}$	$\text{int}_{\text{small}}$	$\text{min}_{\text{small}}$	$\text{int}_{\text{small}}$
$S_1(\pi_1\pi_1{}^*)$	-0.04	-0.01	-0.03	-0.25
$T_2(\pi_2\pi_1{}^*)$	-0.07	-0.03	0.00	-0.04
$T_3(n\pi_1{}^*)$	0.00	-0.03	-0.09	0.05

The second factor that modifies the excitation energies when going from the QM/MM approach to the QM1/QM2/MM approach is the electronic coupling between methylene blue and the surrounding nucleobases. In a multichromophoric system the coupling between transition dipole moments of individual chromophores results in Frenkel exciton states that can be delocalized over more than one monomer of the system.[58,59] In addition, if orbital overlap interactions[60] between monomers is important, exciton or local states can interact with charge-transfer states giving rise to new electronic states with different properties. The parent electronic states are nearly energetically degenerate, and after the mixing they split and the energy of the new formed states can differ significantly from that of the parent states depending on the electronic coupling present in the system. The energy shifts due to the electronic coupling effect, listed in Table 2, are evaluated here as the difference between the CAM-B3LYP excitation energies of the system formed by methylene blue and the surrounding nucleobases (ΔE_{MB+DNA}^{CAM}) and the CAM-B3LYP excitation energies of isolated methylene blue (ΔE_{MB}^{CAM}). As can be seen in Table 2, the energy shifts due to electronic coupling are of the same magnitude (smaller than 0.1 eV) as the polarizing electrostatic contribution except for the $S_1(\pi_1\pi_1{}^*)$ state at the $\text{int}_{\text{small}}$ geometry, which shows a quite pronounced energy variation (-0.25 eV). A closer inspection of this electronic state reveals that the large energy change is a consequence of a strong mixing between the local bright $S_1(\pi_1\pi_1^*)$ state localized in methylene blue and a charge-transfer state between methylene blue and one of the flanking guanines, as is represented in Figure 5. The mixing produces two new electronic states with excitation energies of 2.40 and 2.64 eV and with oscillator strengths of 0.364 and 0.303. The values of the oscillator strengths indicate that the lower-energy state has a larger contribution from the local bright state of methylene blue, and it is the one listed in Table 1, while the higher-energy state has a stronger contribution of the charge-transfer state. The molecular orbitals from which electrons are promoted (π_1 and π_G in Figure 4) during excitation are clearly combinations of molecular orbitals of methylene blue and guanine, showing the mixing character of the two electronic states. The fact that the computed bright state of methylene blue intercalated into $(dG\text{-}dC)_6$ has partial charge-transfer character agrees with the quenching of fluorescence experimentally observed when methylene blue interacts with a polynucleotide $(dGdC)_n$.[36] After Franck-Condon excitation, the system is promoted to the bright state, which presents partial charge-transfer contribution, and then this state likely evolves towards an electronic state with stronger charge-transfer character, and therefore smaller oscillator strength, resulting in a smaller fluorescence quantum yield. This mechanism is similar to what has been proposed to occur in DNA strands, where the initially excited excitons form excimers or exciplexes with charge-transfer character during the early excited-state dynamics.[61]

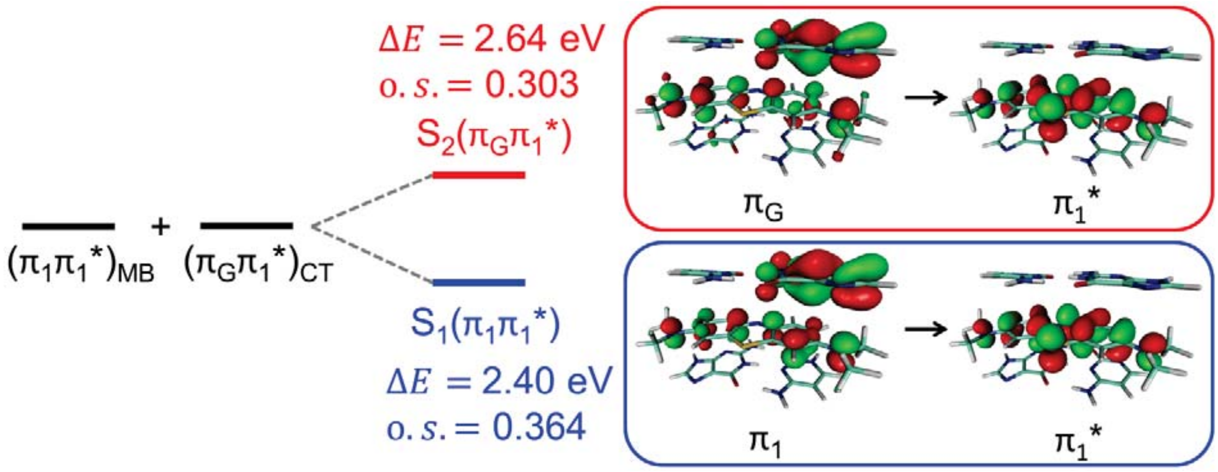


Figure 5. Excitation energies (ΔE), oscillator strengths ($o.s.$) and main electronic transitions of the electronic states $S_1(\pi_1\pi_1^*)$ and $S_2(\pi_G\pi_1^*)$ of methylene blue intercalated into DNA formed after mixing of the local $(\pi_1\pi_1^*)_{MB}$ and charge-transfer $(\pi_G\pi_1^*)_{CT}$ states.

The quality of the results obtained by the two-layered QM/MM and three-layered ONIOM approaches can be evaluated by comparison with experiment. Specifically, the experimental absorption spectrum of methylene blue intercalated into a polynucleotide $(dGdC)_n$ peaks at 674.5 nm (1.84 eV).[36] The computed excitation energies of the bright $S_1(\pi_1\pi_1^*)$ state provided by the QM/MM and ONIOM schemes (Table 1) are 1.98 and 1.72 eV, respectively. Therefore, both approximations display similar errors (0.14 for QM/MM vs 0.12 eV for ONIOM3-EE) and agree very well with the experimental value. Based solely on the excitation energy of the electronic bright state, it is not possible to conclude which approximation is the most accurate one. However, as discussed above, the ONIOM scheme developed here is able to describe electronic coupling between the photosensitizer and the flanking nucleobases, describing properly electronic states of mixing character, which are necessary to explain the quenching of fluorescence experimentally detected.[36] It is worth noting that the influence of the electronic coupling is extrapolated only at the lower quantum mechanical level of theory (in our case TD-

DFT), which means that an energetic correction because of this coupling is added to the excited-state energy computed at the higher level (CASPT2).

3.2 Geometry Optimizations with Quantum Mechanical Environment. In the present section, the impact of describing part of the DNA strand at quantum mechanical level during geometry optimization is analyzed. Two different geometry optimizations were performed using a two-layered electrostatic QM/MM scheme. In the first optimization only methylene blue is in the QM region while in the second geometry optimization both methylene blue and the 4(8) closest nucleobases are in the QM region for the intercalative (minor-groove) structure. The excited-state energies were computed by the ONIOM approach for geometries optimized with the smaller and higher QM regions ($\text{min}_{\text{small}}$ vs $\text{min}_{\text{large}}$ and $\text{int}_{\text{small}}$ vs $\text{int}_{\text{large}}$) and are listed in Table 3. Additionally, the different contributions ΔE_{MB}^{CASPT2}, ΔE_{MB}^{CAM} and ΔE_{MB+DNA}^{CAM} to ΔE^{ONIOM} for all the geometries are shown. The excited-state energy differences ($\Delta E_{large}^{ONIOM} - \Delta E_{small}^{ONIOM}$) between the two geometry optimization procedures are moderate ranging from -0.07 to 0.03 eV. As seen in the previous section, these energy shifts can be explained in terms of electrostatic and electronic coupling effects. The electrostatic effect is caused by the different charges of the nucleobases (force field charges vs B3LYP Mulliken charges) employed during the geometry optimizations, which lead to slightly different geometries. These geometric differences are reflected in different excitation energies. This electrostatic contribution can be computed as the difference between the CASPT2 energies of the geometries optimized using the larger QM region and the CASPT2 energies of the geometries optimized with the smaller QM region, and it is shown in Table 4. As discussed above, the electrostatic effect can increase or decrease the excitation energy of a particular excited state depending on whether the ground state is more or less stabilized than the excited state. For example, the $T_3(n\pi_1^*)$ state is stabilized by 0.05 eV with respect to the ground state when methylene blue is located in the minor groove, while it is destabilized by 0.04 eV when methylene blue is intercalated.

Table 3. ONIOM excited-state energies (in eV) and their contributions ΔE_{MB}^{CASPT2}, ΔE_{MB}^{CAM} and ΔE_{MB+DNA}^{CAM} for methylene blue located into the intercalative and minor-groove pockets of a (dGdC)$_6$ double strand for the two different QM region sizes used for geometry optimization (see text).

	min$_{\mathrm{small}}$				min$_{\mathrm{large}}$				$\Delta E_{large}^{ONIOM} - \Delta E_{small}^{ONIOM}$	$\Delta E_{large}^{ONIOM} - \Delta E^{QM/MM}$
	ΔE^{ONIOM}	ΔE_{MB}^{CASPT2}	ΔE_{MB}^{CAM}	ΔE_{MB+DNA}^{CAM}	ΔE^{ONIOM}	ΔE_{MB}^{CASPT2}	ΔE_{MB}^{CAM}	ΔE_{MB+DNA}^{CAM}		
$S_1(\pi_1\pi_1^*)$	1.92	1.95	2.59	2.56	1.95	1.97	2.60	2.58	0.03	-0.04
$T_2(\pi_2\pi_1^*)$	1.97	1.97	2.04	2.04	1.99	1.99	2.06	2.06	0.02	-0.05
$T_3(n\pi_1^*)$	2.96	3.05	3.12	3.03	2.89	3.00	3.10	2.99	-0.07	-0.16
	int$_{\mathrm{small}}$				int$_{\mathrm{large}}$					
$S_1(\pi_1\pi_1^*)$	1.72	1.97	2.65	2.40	1.68	1.96	2.63	2.35	-0.04	-0.30
$T_2(\pi_2\pi_1^*)$	1.84	1.88	2.02	1.98	1.85	1.89	2.02	1.98	0.01	-0.06
$T_3(n\pi_1^*)$	2.74	2.69	2.72	2.77	2.77	2.73	2.76	2.80	0.03	0.05

The electronic coupling contribution, calculated as $\Delta E_{MB+DNA}^{CAM} - \Delta E_{MB}^{CAM}$, is also shown in Table 4 for the four different optimized geometries together with the difference between the couplings of the large QM region geometries and the small QM region geometries. These electronic coupling differences are only relatively important for the $T_3(n\pi_1^*)$ state of the minor-groove geometry (-0.02 eV) and the $S_1(\pi_1\pi_1^*)$ state of the intercalative geometry (-0.03 eV). In both cases, the energy shift due to electronic coupling is larger for the geometry optimized with the large QM region (-0.11 eV for min$_{\mathrm{large}}$ and -0.28 eV for int$_{\mathrm{large}}$) than for the geometry optimized with the small QM region (-0.09 eV for min$_{\mathrm{small}}$ and -0.25 eV for int$_{\mathrm{small}}$). This indicates that the electronic coupling between the excited states of methylene blue and those of the nucleobases is stronger when the nucleobases are described quantum mechanically during the geometry optimization. This can be seen in Figure 6, where the $\text{min}_{\text{small}}$ and $\text{int}_{\text{small}}$ geometries are compared with the $\text{min}_{\text{large}}$ and $\text{int}_{\text{large}}$ geometries. When methylene blue is located into the minor groove the structure is stabilized by hydrogen bonding between the N atom of the photosensitizer and the NH_2 group of one of the guanines. This hydrogen bond is stronger for $\text{min}_{\text{large}}$ than for $\text{min}_{\text{small}}$ as indicated by the shorter N-H distance and larger N-H-N angle for the $\text{min}_{\text{large}}$ geometry (1.98 Å and 162.3°) than for the $\text{min}_{\text{small}}$ geometry (2.04 Å and 155.8°). When methylene blue is intercalated, the stacking interactions for $\text{int}_{\text{large}}$ are stronger than for $\text{int}_{\text{small}}$. This is reflected in the rise separation between the photosensitizer and the flanking guanine participating in the charge-transfer state that is mixed with the bright state of methylene blue (see discussion in Section 3.1). The rise distance is shorter for the $\text{int}_{\text{large}}$ geometry (3.28 Å) than for the $\text{int}_{\text{small}}$ geometry (3.36 Å). Therefore, for both intercalative and minor-groove binding modes, the interaction between methylene blue and the neighboring nucleobases is stronger when the larger QM region is employed during geometry optimization. This results in larger energetic shifts induced by electronic coupling, as can be seen in Table 4.

Table 4. Polarizing contribution (in eV), computed as the difference between ΔE_{MB}^{CASPT2} for the min$_{\mathrm{large}}$ and int$_{\mathrm{large}}$ geometries and ΔE_{MB}^{CASPT2} for the min$_{\mathrm{small}}$ and int$_{\mathrm{small}}$ geometries, and electronic coupling contribution (in eV), computed as $\Delta E_{MB+DNA}^{CAM} - \Delta E_{MB}^{CAM}$, to the excitation energies ΔE^{ONIOM} of the minor-groove and intercalative geometries. The difference (Diff) of electronic coupling contributions between the large QM region and small QM region geometries is also listed.

	Polarizing Effect		Electronic-Coupling Effect					
	min	int	min$_{\mathrm{small}}$	min$_{\mathrm{large}}$	Diff.	int$_{\mathrm{small}}$	int$_{\mathrm{large}}$	Diff.
$S_1(\pi_1\pi_1^*)$	0.02	-0.01	-0.03	-0.02	0.01	-0.25	-0.28	-0.03
$T_2(\pi_2\pi_1^*)$	0.02	0.01	0.00	0.00	0.00	-0.04	-0.04	0.00
$T_3(n\pi_1^*)$	-0.05	0.04	-0.09	-0.11	-0.02	0.05	0.04	-0.01

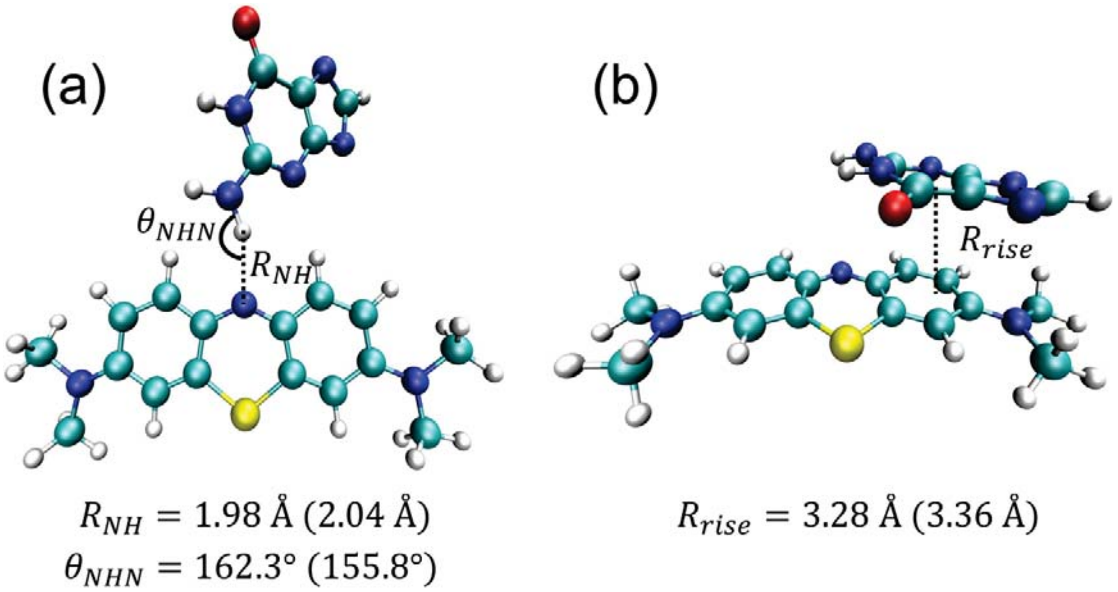


Figure 6. (a) Hydrogen bond N-H distance and N-H-N angle between methylene blue and guanine for the minor-groove structure optimized with the larger QM region. The values for the optimized geometry with the smaller QM region are given in parentheses. (b) Rise separation

between methylene blue and guanine for the intercalative structure optimized with the larger QM region. The values for the optimized geometry with the smaller QM region are given in parentheses.

Finally, we discuss the energy differences between the highest-level and lowest-level calculations (shown in the last column of Table 3): the excited-state energies computed by the ONIOM3-EE approach (ΔE^{ONIOM}) at the geometries optimized with the larger QM region ($\mathrm{min}_{\mathrm{large}}$ and $\mathrm{int}_{\mathrm{large}}$) and the excited-state energies computed by the standard two-layered QM/MM approach ($\Delta E^{QM/MM}$) at the geometries optimized with the smaller QM region ($\mathrm{min}_{\mathrm{small}}$ and $\mathrm{int}_{\mathrm{small}}$). The most significant energy shift is observed for the $T_3(n\pi_1{}^*)$ state at the minor-groove geometry (-0.16 eV) and for the $S_1(\pi_1\pi_1{}^*)$ state at the intercalated geometry (-0.30 eV). The energy shifts for the remaining electronically excited state range between -0.03 and -0.05 eV. As was seen above, these energy differences can be decomposed into the contribution that arises when the environment is described quantum mechanically during the excited-state calculations ($\Delta E^{ONIOM} - \Delta E^{QM/MM}$ column of Table 1) and a second contribution that arises when the environment is described quantum mechanically during the geometry optimization ($\Delta E^{ONIOM}_{large} - \Delta E^{ONIOM}_{small}$ column of Table 3). Both energy contributions are of similar order of magnitude except for the $S_1(\pi_1\pi_1{}^*)$ state, in which the quantum mechanical description of part of DNA during the excited-state calculation (-0.26 eV) is much more important than during geometry optimization (-0.04 eV). In turn, these two contributions can be further decomposed into a polarizing electrostatic effect and electronic coupling effect. The last one, as discussed in Section 3.1, is especially significant for the intercalative geometry, for which the $S_1(\pi_1\pi_1{}^*)$ state undergoes a red shifted of 0.30 eV due to the mixture between the bright state of methylene blue and a charge-transfer state between methylene blue and one of the flanking guanines.

4. CONCLUSIONS

In this manuscript a three-layered electrostatic embedded ONIOM scheme was implemented to perform excited-state calculations on chromophores embedded in biological environments. The developed methodology was applied to calculate the electronically excited-state energies of the photosensitizer methylene blue when it is integrated in the minor groove and the intercalative pocket of a solvated (dG-dC)$_6$ double strand. The ONIOM calculations were compared with standard two-layered electrostatic embedded QM/MM calculations to investigate how different descriptions of the environment influence the excitation energies of the photosensitizer. The ONIOM approach provides a better description of the electrostatic interactions between the different layers and is able to describe electronic coupling between the photosensitizer and the environment.

When part of the environment is described quantum mechanically during the excited-state calculation, excitation energies are converged when 4 and 8 nucleobases are included in the second layer of the ONIOM scheme for the intercalative and minor-groove geometries, respectively. The excitation energies are shifted with respect to energies obtained when the environment is fully described by a force field. Two different factors are involved in the energy shifts, namely electrostatic and electronic coupling components. The electrostatic component is a consequence of the alteration of the environmental charge distribution that interacts with methylene blue. This different charge distribution induced energy shifts up to 0.07 eV in the excitation energies. Although this effect was found to be moderate for methylene blue, it is expected to be more pronounced for other chromophores with more diffuse excitations, e.g.

Rydberg states. Diffuse excited states induce strong alterations on the electronic structure of the nearby environment, and thus a quantum mechanical description is required. The electronic coupling component emerges when an important interaction between the transition dipole moments of the chromophore and the environment exist, generating delocalized exciton states. Additionally, if orbital overlap interactions are also important, local or exciton states can mix with charge-transfer states. The electronic bright state of methylene blue undergoes a strong red shift of 0.26 eV when it is intercalated into DNA. A strong mixing between the local bright state of methylene blue and a charge-transfer state that involves the photosensitizer and one guanine residue is the responsible factor for this large energy change. This effect can only be described if the residues that participate in exciton and/or charge-transfer states are computed at the same level of theory, as it is the case of the second layer of the ONIOM approach, in which methylene blue and the neighboring nucleobases were computed by TD-DFT. Intercalation is a very common binding mode of planar chromophores when they interact with DNA strands. Therefore, it is very likely that electronic coupling plays a relevant role in such situations.

The effect of describing part of the environment at quantum mechanical level during geometry optimizations was also investigated. It was shown that the energy shifts introduced by the electrostatic component were similar as those induced by the quantum mechanical description of the strand during the excited-state calculation. The electronic coupling between methylene blue and the nucleobases is more accentuated when the environment is optimized quantum mechanically. Specifically, it has been shown that hydrogen bonding and stacking interactions are stronger than in the classical description, inducing further energy shift in some of the excitations of the photosensitizer.

AUTHOR INFORMATION

Corresponding Author

*christian.ochsenfeld@cup.uni-muenchen.de

*leticia.gonzalez@univie.ac.at.

Author Contributions

L.G. and C.O. conceived the project. S.R. implemented the ONIOM approach. J.J.N. performed the calculations and analyses. All authors discussed the results and co-wrote the manuscript. J.J.N. and S.R. contributed equally to the manuscript.

Funding Sources

J.J.N. thanks the Marie Curie Actions-Intra-European Fellowships (call FP7-PEOPLE-2012-IEF) for financial support through the project PHOTOBLUE (project number 327412).

S.R. and C.O. acknowledge financial support by the DFG funding initiative SFB749 (TP C7) and the Excellence Cluster EXC114 (CIPSM).

REFERENCES

(1) Brunk, E.; Rothlisberger, U. *Chem. Rev.* **2015**, *115*, 6217-6263.
(2) Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198-1229.
(3) Mennucci, B. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583-6594.
(4) Zobel, J. P.; Nogueira, J. J.; González, L. *J. Phys. Chem. Lett.* **2015**, *6*, 3006-3011.
(5) Barbatti, M. *J. Am. Chem. Soc.* **2014**, *136*, 10246-10249.
(6) Szabla, R.; Šponer, J.; Góra, R. W. *J. Phys. Chem. Lett.* **2015**, *6*, 1467-1471.
(7) Grozema, F. C.; Swart, M.; Zijlstra, R. W. J.; Piet, J. J.; Siebbeles, L. D. A.; Van Duijnen, P. T. *J. Am. Chem. Soc.* **2005**, *127*, 11019-11028.
(8) Parac, M.; Doerr, M.; Marian, C. M.; Thiel, W. *J. Comput. Chem.* **2010**, *31*, 90-106.
(9) Etienne, T.; Very, T.; Perpète, E. A.; Monari, A.; Assfeld, X. *J. Phys. Chem. B* **2013**, *117*, 4973-4980.

(10) Cui, G.; Thiel, W. *J. Phys. Chem. Lett.* **2014**, *5*, 2682-2687.

(11) Nogueira, J. J.; Oppel, M.; González, L. *Angew. Chem. Int. Ed.* **2015**, *54*, 4375-4378.

(12) Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Lewis, J. P.; Gómez-Puertas, P.; Ortega, J. *J. Phys. Chem. Lett.* **2016**, *7*, 4391-4397.

(13) Dumont, E.; Wibowo, M.; Roca-Sanjuán, D.; Garavelli, M.; Assfeld, X.; Monari, A. *J. Phys. Chem. Lett.* **2015**, *6*, 576-580.

(14) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. *J. Chem. Theory Comput.* **2015**, *11*, 4825-4839.

(15) Spata, V. A.; Lee, W.; Matsika, S. *J. Phys. Chem. Lett.* **2016**, *7*, 976-984.

(16) Faraji, S.; Zhong, D.; Dreuw, A. *Angew. Chem. Int. Ed.* **2016**, *55*, 5175-5178.

(17) Lu, Y.; Lan, Z.; Thiel, W. *Angew. Chem. Int. Ed.* **2011**, *50*, 6864-6867.

(18) Zelený, T.; Ruckenbauer, M.; Aquino, A. J. A.; Müller, T.; Lankaš, F.; Dršata, T.; Hase, W. L.; Nachtigallova, D.; Lischka, H. *J. Am. Chem. Soc.* **2012**, *134*, 13662-13669.

(19) Plasser, F.; Aquino, A. J. A.; Hase, W. L.; Lischka, H. *J. Phys. Chem. A* **2012**, *116*, 11151-11160.

(20) Nogueira, J. J.; Corani, A.; El Nahhas, A.; Pezzella, A.; d'Ischia, M.; Gonzállez, L.; Sundström, V. *J. Phys. Chem. Lett.* **2017**, *8*, 1004-1008.

(21) Nogueira, J. J.; Meixner, M.; Bittermann, M.; González, L. *ChemPhotoChem* **2017**, 10.1002/cptc.201600062.

(22) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838-1848.

(23) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11462-11467.

(24) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma Jr, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815-826.

(25) Mellish, K. J.; Cox, R. D.; Vernon, D. I.; Griffiths, J.; Brown, S. B. *Photochem. Photobiol.* **2002**, *75*, 392-397.

(26) Orth, K.; Russ, D.; Beck, G.; Rück, A.; Beger, H. G. *Langenbecks Arch. Surg.* **1998**, *383*, 276-281.

(27) Orth, K.; Rück, A.; Stanescu, A.; Beger, H. G. *The Lancet* **1995**, *345*, 519-520.

(28) Zolfaghari, P. S.; Packer, S.; Singer, M.; Nair, S. P.; Bennett, J.; Street, C.; Wilson, M. *BMC Microbiol.* **2009**, *9:27*.

(29) Ragàs, X.; Dai, T.; Tegos, G. P.; Agut, M.; Nonell, S.; Hamblin, M. R. *Lasers Surg. Med.* **2010**, *42*, 384-390.

(30) Nordén, B.; Tjerneld, F. *Biopolymers* **1982**, *21*, 1713-1734.

(31) Tong, C.; Hu, Z.; Wu, J. *J. Fluoresc.* **2010**, *20*, 261-267.

(32) Tuite, E.; Nordén, B. *J. Am. Chem. Soc.* **1994**, *116*, 7548-7556.

(33) Nogueira, J. J.; González, L. *Biochemistry* **2014**, *53*, 2391-2412.

(34) Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z. *Nucleic Acids Res.* **2005**, *33*, 7048-7057.

(35) Zhang, L. Z.; Tang, G. Q. *J. Photochem. Photobiol. B* **2004**, *74*, 119-125.

(36) Kelly, J. M.; van der Putten, W. J. M.; McConnell, D. J. *Photochem. Photobiol.* **1987**, *45*, 167-175.

(37) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

(38) Hall, K. F.; Vreven, T.; Frisch, M. J.; Bearpark, M. J. *J. Mol. Biol.* **2008**, *383*, 106-121.

(39) Sherwood, P.; De Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J. *J. Mol. Struct. Theochem* **2003**, *632*, 1-28.

(40) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A* **2009**, *113*, 11856-11865.

(41) ChemShell; a Computational Chemistry Shell. www.chemshell.org.

(42) Biancardi, A.; Barnes, J.; Caricato, M. *J. Chem. Phys.* **2016**, *145*, 224109.

(43) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(44) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817-3829.

(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(46) D.A. Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz , A. W.; AMBER 2015, University of California, San Francisco.: 2015.

(47) Crowley, M. F.; Darden, T. A.; Cheatham III, T. E.; Deerfield II, D. W. *J. Supercomput.* **1997**, *11*, 255-278.

(48) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.

(49) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787-1799.

(50) Hariharan, P. C.; Pople, J. A. *Theoret. Chim. Acta* **1973**, *28*, 213-222.

(51) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X. *Mol. Phys.* **2015**, *113*, 184-215.

(52) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663-1669.

(53) Finley, J.; Malmqvist, P. A.; Roos, B. O.; Serrano-Andrés, L. *Chem. Phys. Lett.* **1998**, *288*, 299-306.

(54) Roos, B. O.; Lindh, R.; Malmqvist, P. Å.; Veryazov, V.; Widmark, P. O. *J. Phys. Chem. A* **2005**, *109*, 6575-6579.

(55) Zobel, J. P.; Nogueira, J. J.; González, L. *Chem. Sci.* **2017**, *8*, 1482-1499.

(56) Aquilante, F.; Autschbach, J.; Carlson, R. K.; Chibotaru, L. F.; Delcey, M. G.; De Vico, L.; Fdez. Galván, I.; Ferré, N.; Frutos, L. M.; Gagliardi, L. *J. Comput. Chem.* **2016**, *37*, 506-541.

(57) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51-57.

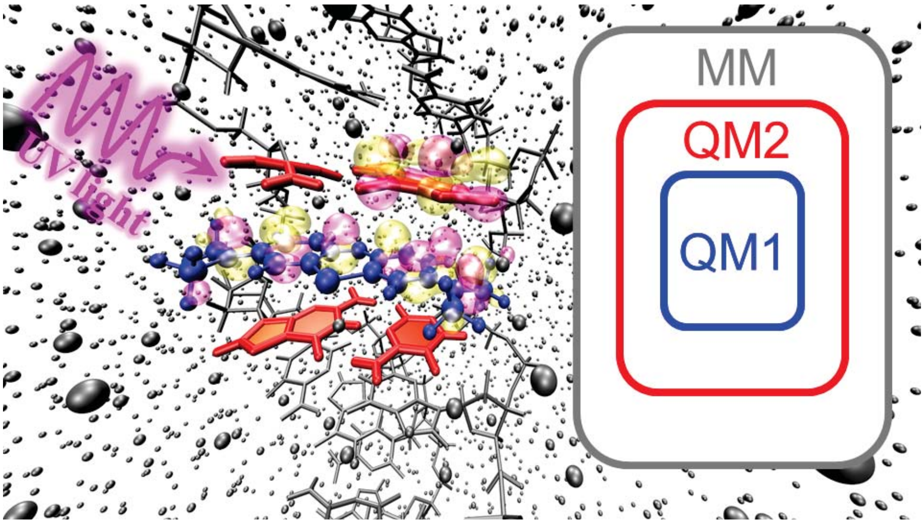
(58) Frenkel, J. *Phys. Rev.* **1931**, *37*, 1276-1294.

(59) Bouvier, B.; Gustavsson, T.; Markovitsi, D.; Millié, P. *Chem. Phys.* **2002**, *275*, 75-92.

(60) Scholes, G. D.; Ghiggino, K. P. *J. Phys. Chem.* **1994**, *98*, 4580-4590.

(61) Schreier, W. J.; Gilch, P.; Zinth, W. In *Annu. Rev. Phys. Chem.* 2015; Vol. 66, p 497-519.

For Table of Contents Only:



31