A EUROPEAN JOURNAL

CHEMPHYSCHEM

OF CHEMICAL PHYSICS AND PHYSICAL CHEMISTRY

Accepted Article

Title: Acidity constant (pKa) calculation of large solvated dye molecules: evaluation of two advanced molecular dynamics methods

Authors: Thierry De Meyer; Bernd Ensing; Sven M. J. Rogge; Karen De Clerck; Evert Jan Meijer; Veronique Van Speybroeck

To be cited as: ChemPhysChem 10.1002/cphc.201600734

Link to VoR: http://dx.doi.org/10.1002/cphc.201600734

A Journal of

ChemPubSoc Europe

www.chemphyschem.org

WILEY-VCH

Acidity constant (pK_a) calculation of large solvated dye molecules: evaluation of two advanced molecular dynamics methods

Thierry De Meyer [†‡] Bernd Ensing [§] Sven M. J. Rogge [†]
Karen De Clerck [‡] Evert Jan Meijer [§*] Veronique Van Speybroeck [†*]

1 Abstract

pH-sensitive dyes are increasingly applied onto polymer substrates for the creation of novel sensor materials. Recently, these dye molecules have been modified to form a covalent bond with the polymer host. This can have a large influence on the pH-sensitive properties, in particular on the acidity constant (pK_a). Obtaining molecular control over the factors that influence the pK_a value is mandatory for future intelligent design of sensor materials. Herein, we show that advanced molecular dynamics (MD) methods have reached the level where pK_a values of large solvated dye molecules can be predicted with high accuracy. Two MD methods are used in this work: steered or restrained MD and the insertion/deletion scheme. Both are first calibrated on a set of phenol derivatives and afterwards applied to the dye molecule Bromothymol Blue. Excellent agreement with experimental values is obtained, which opens perspectives for using these methods for designing dye molecules.

2 Introduction

An acidity constant (pK_a) is a fundamental quantity in solution chemistry. A plethora of reactions, be it chemical or biochemical, are determined by a proton transfer step, where the balance is determined by the pK_a of the compounds participating in the reaction. The reaction studied in this work is the deprotonation of an acid AH to the conjugated base A^- in aqueous solution:

$$AH_{aq} \rightarrow A^-_{aq} + H^+_{aq} \tag{1}$$

of which the pK_a is given by

$$pK_a = -\log K_a = \frac{\Delta G^\circ_a}{k_B T \ln 10} \tag{2}$$

Herein, ΔG°_a is the standard dissociation free energy and k_B the Boltzmann constant. This free energy difference between the protonated and deprotonated state is the central quantity that needs to be calculated.

[†]Center for Molecular Modeling, Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium; E-mail: veronique.vanspeybroeck@ugent.be

[‡]Department of Textiles, Ghent University, Technologiepark 907, 9052 Zwijnaarde, Belgium; E-mail: karen.declerck@ugent.be

[§]Amsterdam Center for Multiscale Modeling and Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands; E-mail: e.j.meijer@uva.nl

[*]corresponding authors

Our specific interest in calculating pK_a values arises from the development of novel sensor materials: halochromic (pH-sensitive) dyes are incorporated into polymeric structures, resulting in pH-sensitive polymers which can be used in wound bandages, protective clothing, etc.[1, 2, 3, 4, 5, 6, 7, 8, 9, 10] In recent developments, these dye molecules are modified to include reactive or polymerisable groups.[11] This results in a material where the dye is covalently bound to the host material, greatly improving the leaching properties. This modification, however, changes the molecular structure of the dye and therefore also its halochromic properties. For a pH-sensitive wound bandage, for example, a pK_a around $6.5 - 7.0$ would be ideal.[12] The synthesis and purification of these modified dye molecules can be a tiresome process and theoretical calculations can provide a true added value for obtaining molecular insight into the factors governing the pH-sensitive behaviour. Ideally, dye molecules could be designed to yield halochromic behaviour in the desired pH-region. This requires techniques which are not only able to explain experimentally observed pH-sensitive properties, but are accurate enough to be used in a predictive manner.

Because of the importance of acidity constants in solution chemistry, a lot of effort has been put into the calculation of these values.[13, 14, 15] Thanks to developments in solvent reaction field methods,[16, 17, 18] where the solute is placed inside a cavity in a dielectric continuum, pK_a's of molecules in solution have been successfully calculated using static calculations.[19] This procedure relies on the calculation of a limited number of points on the potential energy surface. Mostly, a thermodynamic cycle is used and the only empirical input is the aqueous solvation free energy of the proton.[20, 21, 22, 23, 24, 25, 26] It has been shown that adding explicit solvent molecules inside the cavity can improve the accuracy of calculated pK_a's, although it is challenging to determine the necessary amount of water molecules.[27, 28, 29, 30, 31, 32, 33]

To treat more complex molecular environments, such as polymeric materials, one needs to go beyond continuum techniques. A more advanced approach is to explicitly take the solvent into account, treating both the compound and environment (in this case the solvent) at the same level of theory. This can be done by employing density functional theory (DFT) based molecular dynamics simulations (DFTMD).[34, 35] The advantage of DFTMD is the uniform treatment of the entire system, allowing not only electronic polarisation of the solvent and solute, but also explicit interactions such as hydrogen bonds. Herein, we use such techniques since they allow to sample the entire proton transfer process while taking the dynamic solvent environment into account.[36, 37]

Previously we were able to experimentally measure the pK_a values of a range of sulfonphthaleine dyes, for which the general structure is visualised in Figure 1.[38, 39] The next step in designing dyes engineered towards specific applications, however, requires a thorough understanding of the molecular factors governing the pK_a. A reliable procedure is needed to reach this goal. Hereto, we use two advanced MD methods that fully capture the effects of the molecular environment, which are first evaluated on their validity for a set of phenol derivatives. Computational screening of several sulfonphthaleine dyes would be too computationally demanding. The phenol derivatives chosen in this work bear substituents similar to those found in sulfonphthaleine dyes (Table 1) and were previously studied by static approaches.[21] Afterwards, the most suited procedure will be applied to Bromothymol Blue, a sulfonphthaleine dye for which R_1=*i*-Pr, R_2=Br and R_3=CH_3 and for which we previously found a pK_a of 7.4.[38]

This paper is organised as follows. First we give a brief overview of the two MD techniques for the determination of pK_a values, after which the results for the substituted phenols and Bromothymol Blue are presented.

3 Advanced MD techniques for pK_a determination

Proton transfer reactions have been the subject of various studies.[40, 41, 42, 43] The difficulty in studying these reactions with MD methods is that a proton transfer is a rare event. Advanced MD methods are, therefore, required to simulate these processes, as the time scale of a MD run is typically of the order of a few ps. One class of methods applies an external biasing potential to transfer the proton in a controlled manner along a reaction coordinate. This is often referred

to as restrained or steered MD and was pioneered by Sprik and co-workers specifically for pK_a calculations.[44, 45, 46, 47, 48, 49, 50, 51, 52] The idea behind this method is to integrate the force needed to "drag" the system from the protonated to the deprotonated state, from which the free energy can be determined.[53] The choice of reaction coordinate can sometimes be cumbersome as will be explained in Section 3.1 and Appendix A.1.

The group of Sprik developed another interesting approach to calculate pK_a's.[54, 55, 56, 57] This method concerns the reversible elimination of protons and electrons, allowing to calculate redox potentials and acidity constants based on the principles of Marcus' theory of electron transfer.[58, 59] The pK_a calculation itself is based on an insertion/deletion scheme, where reaction (1) is split into two half reactions and the free energy change can be determined based on thermodynamic integration. This method has been successfully applied to several aqueous systems and Cheng et al. concluded that an accuracy of 1–2 pK_a units is feasible.[60, 61, 57] This method is discussed in Section 3.2 and Appendix A.2.

It was shown previously that performing MD simulations can greatly improve the accuracy of calculated absorption wavelengths of dye molecules (in combination with TD-DFT) compared to static calculations[62] As will be shown here, MD simulations are also effective for calculating pK_as of large solvated systems. In Section 3 the theory of restrained MD and the insertion/deletion method will be discussed. In Section 4 both methods are applied to the phenol derivatives and the results will be discussed. Afterwards, the solvated dye Bromothymol Blue will be studied.

3.1 Restrained/steered MD

During a restrained MD simulation,[63, 64] the system is gradually moved along a chosen reaction coordinate (collective variable) from the reactant state (e.g., the acid AH_{aq}) to the product state (e.g., $A^-_{aq} + H^+_{aq}$). The free energy difference between reactant and product state is equal to the work required to shift the system along the chosen coordinate, which can be calculated by integrating the force along the coordinate:[65, 53]

$$k_B T \ln 10 \ \mathrm{pK}_a = \Delta F = -\int_{q_0}^{q_1} \langle f \rangle_q \, \mathrm{d}q \tag{3}$$

in which we set $F(q_0) = 0$. q needs to be a well-chosen reaction coordinate, in which q_0 and q_1 correspond to the reactant and product state, respectively. This integral is approximated by keeping the system fixed at several values of q by a strong spring, while still allowing for small fluctuations. This is done through the software package PLUMED.[66, 67] From a thermodynamic point of view, Hess' law ensures that the resulting pKa is independent of the chosen reaction path. However, from a computational point of view, the description of the molecular reaction depends on the choice of one or multiple collective variables. If poorly chosen, the subsequent sampling of the phase space may be influenced by this choice. Therefore, the specific choice of the coordinate space is crucial to sample the relevant parts of the free energy surface.[68, 69, 70, 71] Two coordinates are commonly used to study proton transfer reactions: a coordination number or a difference in distance.[53] These coordinates have the issue that they are not able to track the position of the acid proton during the entire deprotonation reaction, which is explained in detail in Appendix A.1. To circumvent this issue, a different approach is proposed here. In the complete reaction the proton is transferred from its position in the parent molecule (bonded to the donor oxygen) to the aqueous solution. We will assume that the latter occurs in the form of a hydronium ion, but this will be discussed further on. As was observed in simulations with above-mentioned collective variables,[53] the acid proton escapes and migrates further away from the donating atom into the solvent box. The product state of the deprotonation reaction is indeed the one where the anion and hydronium are as far away from each other as possible, allowing both ions to have their own solvation layer. During this reaction, the coordination number of the donating oxygen atom will drop and the coordination number of the water molecule that accepts the acid proton will increase. Therefore, we propose to use a difference in coordination number as the reaction coordinate:

$$\begin{aligned}\Delta n_c &= n_c(\mathrm{O_{AH}}) - n_c(\mathrm{O}_w) \\ &= \sum_i \frac{1-\left(\frac{r(\mathrm{O_{AH}}-\mathrm{H}_i)}{r_0}\right)^n}{1-\left(\frac{r(\mathrm{O_{AH}}-\mathrm{H}_i)}{r_0}\right)^m} - \sum_i \frac{1-\left(\frac{r(\mathrm{O}_w-\mathrm{H}_i)}{r_0}\right)^n}{1-\left(\frac{r(\mathrm{O}_w-\mathrm{H}_i)}{r_0}\right)^m}\end{aligned} \tag{4}$$

The choice of the accepting water molecule will be discussed in Section 4.1. For a water molecule in the second solvation layer, this coordinate is illustrated in Figure 2.

The goal of using this coordinate is to be able to sample the entire reaction: from the acid molecule in solution to the deprotonated state, including the migration of the proton further away from the conjugated base. This also includes the sampling of a proton wire when the proton is transferred through the solution. The cutoff radius r_0 is set to 1.2 Å. At intermediate values of Δn_c, this will correspond to a situation where both the donating and accepting oxygens tend to have a proton in their proximity. The acid proton will therefore remain between both oxygens, putting some strain on the proton wire. The smoothing parameters n and m were chosen to be respectively 8 and 16, resulting in a relatively steep switching function. It is also important to mention that for 4-hydroxybenzoic acid, a quadratic wall was used to keep the coordination number of both oxygens of the carboxylate group low, otherwise the acid proton would simply reattach to the phenol at the carboxylate group instead of remaining in the solvent.

To summarize, the difference in coordination numbers Δn_c allows us to effectively sample the entire reaction, including the proton wire, in a one dimensional coordinate.

3.2 Insertion/deletion

The particle insertion/deletion scheme was first proposed by the Sprik group.[54, 55, 57] The essential parts of the methodology are summarised here; for a more in-depth discussion the reader is referred to the literature of the Sprik group.[56, 54, 55, 57]

The total reaction (1) is split up into two half reactions:

$$\mathrm{AH_{aq}} \rightarrow \mathrm{A^-_{aq}} + \mathrm{H^+_g} \tag{5}$$

$$\mathrm{H^+_{aq}} \rightarrow \mathrm{H^+_g} \tag{6}$$

Herein, an intermediate step in the gas phase (subscript g) is used. One can interpret this as the proton being removed from the solvated acid AH to the gas phase (deletion) and inserted in a pure water box, as illustrated in Figure 3. By combination of the two simulations, the total free energy change can be obtained. The derivation followed here is written down for the acid AH, but is identical for a proton in solution (for instance in the form of a hydronium ion, $\mathrm{H_3O^+}$).

The total energy E is defined as a linear function depending on a coupling parameter η, which can take values from 0 to 1:

$$E_\eta(\eta) = (1-\eta)E_{\mathrm{AH}} + \eta E_{\mathrm{A^-}} \tag{7}$$

For $\eta = 0$, the system is in the protonated state (i.e. AH), while for $\eta = 1$ the deprotonated state ($\mathrm{A^-}$) is retrieved. The intermediate values correspond to a mixture of AH and $\mathrm{A^-}$, meaning a partial deprotonation, which has no experimental counterpart.

The simulations in this work will be performed in the NVT ensemble, which implies constant volume V. The pK_a will therefore be calculated from a Helmholtz free energy change ΔF instead of a Gibbs free energy change ΔG. The difference in free energy, ΔF, can then be written down as

$$\Delta F_\eta = F_\eta(1) - F_\eta(0) = \int_0^1 \frac{\partial F(\eta)}{\partial \eta} \mathrm{d}\eta \tag{8}$$

To determine the derivative of the free energy with respect to the coupling parameter, the partition function of a system with an energy function that depends on η is first written down:

$$Q(N,V,T,\eta) = \frac{1}{\Lambda^{3N}N!}\int \mathrm{d}\boldsymbol{r}^N \exp[-\beta E_\eta(\eta)] \tag{9}$$

with N the number of particles, Λ the thermal wavelength and $\beta = (k_B T)^{-1}$.

The free energy derivative can then be written as an ensemble average:

$$\begin{aligned}\left.\frac{\partial F(\eta)}{\partial \eta}\right|_{NVT} &= -\frac{1}{\beta}\frac{\partial}{\partial \eta}\ln Q(N,V,T,\eta) \\ &= -\frac{1}{\beta Q(N,V,T,\eta)}\frac{\partial Q(N,V,T,\eta)}{\partial \eta} \\ &= \frac{\int \mathrm{d}\boldsymbol{r}^N (\partial E_\eta(\eta)/\partial\eta)\exp[-\beta E_\eta(\eta)]}{\int \mathrm{d}\boldsymbol{r}^N \exp[-\beta E_\eta(\eta)]} \\ &= \left\langle \frac{\partial E_\eta(\eta)}{\partial \eta}\right\rangle_\eta\end{aligned} \tag{10}$$

Employing definition (7), the derivative of the internal energy can be written as

$$\left\langle \frac{\partial E_\eta(\eta)}{\partial \eta}\right\rangle_\eta = \langle \Delta E\rangle_\eta = \langle E_{\mathrm{A}^-} - E_{\mathrm{AH}}\rangle_\eta \tag{11}$$

This formula requires careful interpretation. Although ΔE is defined as a vertical energy gap and in that way independent of η, there is still an implicit dependence of η. Indeed the potential energy surface, and thus the nuclear coordinates, are dependent on the specific value of η. For the propagation of the MD simulations, the forces are weighted by η, as is illustrated in the workflow given in Figure 4. Each value of η defines a different mixing and therefore also requires a different MD simulation. As a result one obtains accumulative averaging energy gaps as a function of η (vide infra).

Finally, the free energy difference (8) can be written as

$$\Delta F_\eta = \int_0^1 \frac{\partial F(\eta)}{\partial \eta}\mathrm{d}\eta = \int_0^1 \langle \Delta E\rangle_\eta \,\mathrm{d}\eta \tag{12}$$

This means that the free energy difference can be directly calculated from the average vertical energy difference $\langle \Delta E\rangle_\eta$ between AH and A^- during simulations at various values of η.

In a linear approximation, the thermodynamic integral (12) can be obtained as:

$$\Delta F = \frac{1}{2}(\langle \Delta E\rangle_0 + \langle \Delta E\rangle_1) = \frac{1}{2}(\Delta E_{\mathrm{AH}} + \Delta E_{A-}) \tag{13}$$

where $\Delta E_{\mathrm{AH}} = \langle \Delta E\rangle_0$ is the ensemble average of the vertical deprotonation energy gap of the acid AH ($\eta = 0$) and $\Delta E_{\mathrm{A}^-} = \langle \Delta E\rangle_1$ is the corresponding average for the conjugated base A^- ($\eta = 1$).

When the proton is inserted/removed, this will, in general, give rise to a substantial reorganisation of the surrounding solvent molecules. Non-linear effects are therefore expected and the numerical integration of (12) may require a better quadrature using a three-point integration formula (evaluating in $\eta = 0$, 0.5 and 1):

$$\Delta F = \frac{1}{6}(\langle \Delta E\rangle_0 + \langle \Delta E\rangle_1) + \frac{2}{3}\langle \Delta E\rangle_{0.5} \tag{14}$$

Instead of the three-point trapezium quadrature (Simpson's rule), a Gauss-Legendre quadrature can also be used:

$$\Delta F = \frac{5}{18}(\langle \Delta E\rangle_{0.1127} + \langle \Delta E\rangle_{0.8873}) + \frac{4}{9}\langle \Delta E\rangle_{0.5} \tag{15}$$

Besides providing accuracy up to higher order,[72, 73] the latter has the advantage that the end points of the integration interval are not taken into account (as explained in Section 4.2). The pratical implementation and final pK_a expression are discussed in Appendix A.2.

4 Results and Discussion

Table 1 lists the molecules under study with their experimental pK_as. For each molecule, the acidity constants will be calculated and rationalised. The pK_a of the different phenol derivatives can be easily understood from the electron donating/withdrawing effect of their substituent(s).[38] The halogens in compounds **2**−**7** cause the pK_a to drop with respect to the non-functionalised phenol (**1**). The electron withdrawing effect of these substituents results in decreased charge density in the aromatic system, meaning a more acidic environment is needed to stabilise the acid proton. A methyl group has a small electron donating effect, which translates in a slightly higher pK_a for compounds **8** and **9**. Compound **10** has a slightly lower pK_a than **1**, indicating that the carboxylate group, even though it is already negatively charged, still allows for some stabilisation of the extra negative charge upon deprotonation of **10**. For the chlorine substituted phenols, **2**−**4**, a clear trend of the substituent as a function of its position can also be seen. How accurate these effects can be reproduced will depend on how well the chosen level of theory is able to capture these electron donating/withdrawing effects.

4.1 Restrained MD

In this work, a difference in coordination number (Δn_c) is used as reaction coordinate, as explained in the previous section. Δn_c is calculated between the donating oxygen of the phenol (the acid AH) and the oxygen of the accepting water molecule. The most important choice to be made is the selection of the water molecule the proton is transferred to, hence becoming the hydronium ion. Phenol was taken as a reference system and three different water molecules were taken as candidates to become the hydronium. The first choice was obvious: the −OH group of phenol has a hydrogen bond with one water molecule (**A** in Figure 5). This water molecule has a hydrogen bond with a molecule in the second solvation layer, making this molecule our second choice (**B**). Following the hydrogen bond network, a third water molecule (**C**) was also chosen. The larger the distance between the phenol anion and the hydronium, the more stable each ion can be in their own solvation layer. From the three selected molecules, however, simulations where we selected molecule **C** were unstable. The acid proton did not remain between phenol and the chosen water molecule and escaped to the surrounding solvent. This means that the collective variable as constructed here was not able to maintain the proton wire. For water molecules **A** and **B**, however, the simulations were stable.

Two different simulations were then performed; with Δn_c defined as the difference between the coordination number of the phenol oxygen (O_{Ph}) and the oxygen of water molecule **A** (O_A) in one simulation and with Δn_c between O_{Ph} and the oxygen of water molecule **B** (O_B) in the other. Δn_c was varied between −0.836 and −2.177. These values correspond with the protonated and deprotonated phenol and were estimated by averaging coordination numbers in short MD runs. From a chemical point of view, varying Δn_c from −0.836 to −2.177 can be interpreted as the donating oxygen atom repelling neighbouring protons, while the accepting oxygen atom will attract a third proton.

For each water molecule, snapshots corresponding to three values of Δn_c are given in Figure 6. These were taken at the end of the simulations, so after approximately 35 ps. For **A**, one can clearly see that the proton hops from the phenol to the neighbouring water molecule. For **B**, the proton first hops to the closest water molecule. All molecules are rather close to each other, corresponding with a proton wire. In the final snapshot, the reaction is completed and the hydronium and phenol anion move away from each other. This is not observed in situation **A**, which can be understood when looking at the corresponding changes in free energy (Figure 7).

For **A**, the free energy keeps rising as a function of Δn_c: there is no clear plateau, which makes it difficult to define a point where the "stable" product is formed. Ions have the tendency to diffuse away from eachother in aqueous solution, which is not possible with Δn_c as constructed for **A**, explaining why no plateau is formed. For **B**, however, a plateau is reached between the 8^{th} and 9^{th} point, as the free energy barely changes, which corresponds with a stable state. The absence of a clear product valley in the free energy profile, while the reaction has clearly finished (as can be seen from Figure 6(f)), may be explained as follows. Phenol is an alkalic compound and thus the most stable state will always correspond to the one where the proton is bound to the phenol. It is expected that in larger solvated systems the hydronium ion and anion would diffuse further away from each other, which would be entropically favoured. In the current set up of the system with a solvated box of 64 water molecules it would be impossible to simulate such effects. For further simulations using the restrained MD method we used the simulation where **B** was selected as accepting water molecule and chose the energy corresponding to the plateau to deduce the free energy difference. In all simulations, this state corresponds to a slightly lower coordination number of the accepting oxygen atom (O_B). This allows the hydronium ion to reside in Zundel or Eigen ionic structures, which were indeed shown to be lower in free energy.[40]

Using the full free energy profile of the simulation where **A** is chosen as the hydronium, ΔF is 50.5 ± 4.1 kJ/mol, which corresponds to a pK_a of 8.9 ± 0.7. For the simulation where **B** is the hydronium, ΔF is 54.7 ± 4.0 kJ/mol when taking into account points 1 to 9 (and omitting the final point), which corresponds to a pK_a of 9.7 ± 0.7. These two values are close to each other and both are reasonable estimates of the experimental pK_a of 10.0. As discussed before, only for the simulation with **B** as the hydronium there is a clearly observable plateau, making the end point for the integration interval non-arbitrary. Therefore, for the calculations on all phenol derivatives, a water molecule from the second solvation layer is chosen. The results of these simulations are given in Table 2; energetic values can be found in the Supporting Information (SI).

Overall, the results obtained with the restrained MD simulations reproduce quite reliably the experimentally observed trend in the acidity constants. Some simulations, however, showed the same instability as was observed with molecule **C** in the case of phenol. More specifically for molecules **3**, **4** and **7** the proton escaped to the solvent when simulating at $\Delta n_c = -1.73$. Molecule **6** proved more problematic: the simulation was also unstable at $\Delta n_c = -1.43$ and -1.88. As mentioned in Section 3, the proton wire is only stable if both oxygens (between which Δn_c is calculated) still want the acid proton in their proximity. These compounds have a lower pK_a than phenol, meaning they will have a lower inclination to have the proton closeby. These points correspond to the situation where the phenol is completely deprotonated but where the chosen water molecule can not become a hydronium yet (due to the chosen value of Δn_c). For molecule **6**, both bromine substituents also give sterical hindrance for the formation of a proton wire, making it extra difficult to sample the reaction as performed here. Therefore, the above mentioned points are omitted when calculating the pK_a's and it is stressed that only stable simulations were used for the final results. Nevertheless, the results obtained with restrained MD method are in good agreement with experiment for all phenols: the root mean square deviation is only 0.5 pK_a units, with an average standard deviation of 0.9. The deviation from experiment is very low since 2.5 pK_a units is generally considered as chemically accurate.[14]

Even though the results are in excellent agreement with experimental values, the use of a difference in coordination number has two downsides. The first is general to all restrained MD simulations: the limited box size. As already mentioned before, the phenol anion and hydronium are most stable when they can be far away from each other; allowing each to have its own solvation shell. Secondly, the simulation was found to be unstable at some points of Δn_c as the proton did not remain in the originally defined proton wire, but rather diffused away. These shortcomings will be addressed in the following section by employing the insertion/deletion method.

4.2 Insertion/deletion

In the insertion/deletion scheme, the deprotonation of the acid (phenol or dye molecule) and protonation of the hydronium ion are considered in different simulations (schematically represented

in the thermodynamic cycle of Figure 3). One could consider both to be at an infinite distance from each other. Furthermore, the proton "disappears" from the simulation into the gas phase, making it unnecessary to sample a proton wire as was the case with restrained MD. Therefore, the insertion/deletion scheme can address both shortcomings found previously.

The two quadratures mentioned in Section 3.2 to estimate the thermodynamic integral (12) are evaluated here. For the trapezium rule (14), the average energy difference $\langle \Delta E \rangle$ between the protonated and deprotonated system is evaluated during simulations of η equal to 0, 0.5 and 1. For $\eta = 1$, the dummy atom is a non-interacting particle. Since it is essentially absent, the surrounding water molecules will arrange around the negatively charged oxygen atom and form hydrogen bonds. At each point of the simulation, however, ΔE is calculated between the dummy equal to a non-interacting particle and equal to a proton. This might lead to energetically very unfavourable conformations, which can be seen from Fig. 8: for the simulation at $\eta = 1$ strong negative peaks can be seen throughout the simulation. This is easy to understand from a chemical point of view: these peaks correspond to geometries where the dummy atom is very close to the proton of a water molecule, which forms a hydrogen bond with the negatively charged oxygen atom.

The Gauss-Legendre quadrature (15) has the advantage that the end points of the integration interval are not taken into account. More specifically in our case, $\langle \Delta E \rangle$ is evaluated during simulations of η equal to 0.1127, 0.5 and 0.8873. Chemically, this means that the proton is never fully absent: for the highest value of η, the dummy atom is still partially a proton. Figure 8 shows that the strong negative peaks in ΔE are gone for the simulation at $\eta = 0.8873$. This will lead to faster convergence of the integration and therefore the Gauss-Legendre quadrature will be used further on. Figure 9 shows the values and running averages of ΔE for phenol for the three values of η. It is clear that very large fluctuations in ΔE are observed, even with the Gauss-Legendre quadrature. It is therefore important to perform relatively long MD simulations to obtain statistically relevant results.

The insertion/deletion method as discussed above was then applied to all phenols and the results are given in Table 2. The calculated $pK_{a,i/d}$'s are in good agreement with the experimental ones. The root-mean-square difference between experiment and theory is 0.9 pK_a units. Including ZPE corrections can have a large effect on the final pK_a: its effect can be negligible or mount up to 0.8 pK_a units. On average, the effect is certainly non-negligible and lowers the root-mean-square deviation to 0.7 pK_a units. With inclusion of the ZPE corrections, the deviation is comparable to the results found with restrained MD. The deprotonation energies with standard deviations can be found in the SI. The standard deviation on the pK_a's is on average 1.0. Summarising, the insertion/deletion scheme is capable of predicting pK_a values in good agreement with experiment.

The goal of this contribution is to apply advanced MD methods for the calculation of pK_a values for large solvated dye molecules. For reference, a static approach has been evaluated too and is discussed in the SI. Ho and Coote recommended usage of computationally expensive methods such as G3MP2 or CBS-QB3 for the gas phase energies.[33] For Bromothymol Blue, an estimation based on DFT is made in the SI, as such methods are not feasible and beyond the scope of this study. In the previous part, the restrained MD method was shown to provide accurate results compared to experiment, but the insertion/deletion scheme provided the most stable simulations, as certain shortcomings found with restrained MD could be circumvented. Therefore, preference was given to this method for application to the dye molecule Bromothymol Blue.

A snapshot during the MD simulation of the solvated Bromothymol Blue system is shown in Figure 10. The system consists out of 473 atoms in total, which is very large and explains the slight adaptation in computational parameters (see Section 6). The pK_a calculated through the insertion/deletion scheme amounts to 7.8 ± 1.1, which is very close to the experimental value of 7.4. The ZPE corrections were also calculated, but in this case had no effect on the final pK_a.

Sampling at different values of η has, of course, a large influence on the solvation of the solute, which is illustrated in Figure 11. Herein, the radial distribution functions (RDF) for solvated phenol (compound **1**), as prototype for the phenol derivatives, and Bromothymol Blue are given. In this case, the RDF is a probability distribution indicating the probability of finding a solvent proton at a certain distance from the donating oxygen. To determine the amount of protons at a

certain distance (i.e., in a specific solvation layer), the RDF can be integrated. This results in the cumulative RDF (CRDF), which is also shown in Figure 11. The acid proton (the dummy atom) is not included in the RDF calculation, as this would give a constant peak around 1 Å, regardless of the value of η.

For the phenol molecule (Figure 11(a)), the following observations were made. With $\eta = 0.1127$, the RDF shows a relatively small peak around 1.87 Å. This peak defines the position of the first solvation layer around the phenol oxygen, which corresponds to a hydrogen bond. The averaged amount of water molecules in this layer can be determined from the value of the CRDF after the first solvation layer, which is about 0.75. This can easily be understood: the phenol oxygen is still bonded to the dummy atom (which is almost a full proton) and therefore has only one free electron pair for accepting a hydrogen bond. Apparently, 75 % of the time, such a hydrogen bond is indeed present. For $\eta = 0.8873$, the dummy atom is almost completely non-interacting (without a charge). The first solvation layer is much closer, around 1.71 Å, and from the CRDF it is derived that on average 2.3 water molecules are located in this solvation layer. The simulations for $\eta = 0.50$ show intermediate values, with a first solvation layer at 1.77 Å which contains on average 1.4 water molecules. These results confirm that we are effectively sampling a partial proton and also clearly show that the solvent undergoes large conformational changes; a three-point quadrature (instead of a linear approximation) is therefore indeed necessary.

For Bromothymol Blue (Figure 11(b)), a very similar effect is seen on the RDF. The largest difference with the RDFs for phenol is the height of the peaks. This can be attributed to sterical hindrance around the acid OH group in the dye molecule (see Figure 1). Specifically for $\eta = 0.1127$, the probability of finding a solvent proton nearby is almost negligible. For $\eta = 0.50$, a first solvation layer is observed at 1.93 Å, which contains on average 0.6 water molecules. When the acid proton is almost completely removed ($\eta = 0.8873$), a much stronger solvation is observed. The first solvation layer is shifted to 1.81 Åwith on average 1.84 water molecules. Notice this is still less than observed for phenol (2.3 water molecules), which again shows the effect of sterical hindrance.

The employed MD simulations combined with the insertion/deletion scheme are thus accurate enough to be used in a predictive manner for complex, realistic systems. The ability to predict acidity constants is a big step towards the development of novel sensor materials. More specifically, this allows to predict the pK_a of modified dye molecules or to study the effect of different, more complex environments.

5 Conclusions

For the design of intelligent materials, it is crucial to be able to govern the factors determining their sensitivity. In the case of pH-sensitive polymers, the acidity constant (pK_a) is of utmost importance, as it determines the pH-region where the sensor is active. In this contribution, we showed that advanced molecular dynamics (MD) are capable of calculating pK_a values for large solvated dye molecules.

Two molecular dynamics (MD) based methods were discussed and first evaluated on a set of phenol derivatives. The first method is restrained MD, where a novel reaction coordinate is proposed. The commonly used coordinates, namely a coordination number or a difference in distance, show difficulties in maintaining control of the entire reaction. Therefore, a different coordinate is utilised, namely a difference of coordination numbers Δn_c. This coordinate allows us to sample the full deprotonation reaction, including a proton wire. It was observed that the free energies formed a plateau at values of the coordinate corresponding with Zundel or Eigen ionic structures. The average deviation from experiment was only 0.5 pK_a units, making this approach very promising.

The second method is the insertion/deletion scheme, as originally proposed by the group of Sprik. This procedure has the advantage that the deprotonation of the phenol and protonation of a water molecule (to a hydronium) are sampled in different simulation boxes. No issues with stability of the simulation arise and the phenol and hydronium ion can be considered to be infinitely far way from each other. The results obtained with this method were also in good agreement with

experiment, with an average deviation of 0.7 pK_a units.

Even though both methods provided good results compared to experiment for the phenol derivatives, the insertion/deletion scheme did not show the same issues as the restrained MD simulations (such as the escaping proton). Therefore, this method was chosen to apply to the sulfonphthaleine dye Bromothymol Blue. The deviation between experiment and theory was found to be only 0.4 pK_a units, confirming the insertion/deletion scheme as a reliable method for calculating pK_a's of large solvated dye systems. This procedure illustrates that the presented computational techniques are accurate enough to predict acidity constants of these large, complex systems. This will allow us to predict pK_a's of modified dyes or dyes in more complex environments in the future.

6 Computational details

All simulations in this work are performed using the CP2K/Quickstep package.[74] Electronic structures are calculated with density functional theory, which is implemented based on a hybrid Gaussian plane wave (GPW) approach.[75] The BLYP functional is used for the exchange-correlation,[76, 77] while Grimme D3 dispersion corrections are used to improve van der Waals interactions.[78] GTH pseudopotentials are employed to avoid calculations of core configurations.[79] The GTH-TZV2P basis set was used for the Gaussian basis and the plane wave kinetic energy cutoff was set to 400 Ry. For bromine, the MOLOPT DZVP basis set was chosen.[80]

To allow for energy conservation during the Born-Oppenheimer molecular dynamics, the wave function optimisation tolerance was set to $0.3\,10^{-7}$ Hartree. To enlarege the MD time step, we replaced the hydrogens atoms by deuterium atoms, allowing to maintain energy conservation with a 0.75 fs time step. Temperature was initially brought to the desired 320 K with a CSVR thermostat, after which temperature was maintained at 320 K using a Nosé-Hoover chain thermostat with 4 beads.[81, 82] The higher temperature was chosen to avoid the *glassy* (over structured) behaviour of BLYP liquid water found at lower temperatures.[83]

The models used in this work consist of 3D periodically repeated cubic cells with a lattice parameter of 12.462 Å. This corresponds to the volume of 64 water molecules at experimental density, which is used for the pure water simulations. For the solvated phenols, the same system size is used. This allows us to rely on error compensation when calculating the total free energy change for the insertion/deletion scheme. From the solvated phenol systems, an amount of water molecules were removed to maintain atmospheric pressure. This amount was estimated by performing serveral force field simulations at fixed volume with a different amount of water molecules and averaging the pressure, using a pure water simulation as reference. The force fields were generated by antechamber (part of AmberTools) and made use of the General AMBER Force Field (GAFF).[84, 85] For the solvated phenol, 59 water molecules were required to achieve atmospheric pressure, while for the singly and doubly substituted components 58 and 57 water molecules were needed respectively. 4-hydroxybenzoic acid required 59 water molecules, as the carboxylate group strongly interacts with the solvent.

Bromothymol Blue required a larger solvent box, namely a cubic cell with a 16.863 Å lattice size, corresponding to the volume of 160 water molecules at experimental density. To obtain atmospheric pressure, 137 water molecules were needed to solvate Bromothymol Blue. Due to the system size, the SCF convergence setting was lowered to $1\,10^{-6}$ Hartree; combined with a plane wave kinetic energy cutoff of 320 Ry and a GTH-TZVP basis set. The same settings were also chosen for the pure water simulation with 160 molecules.

Initial structures of the solvated systems were generated by Packmol.[86] Radial distribution functions were calculated through the analysis packages in YAFF;[87] visualisation of the MD simulations was done with VMD.[88] Block averaging methods were used to estimate standard deviations on the obtained forces and $\langle \Delta E \rangle$'s.[89] 3D molecular representation were generated by CYLview.[90]

For the restrained MD simulations, each simulation point was equilibrated for at least 12 ps, after which the forces were averaged over 19 ps (25 000 steps). For the insertion/deletion scheme,

the equilibration runs were minimal 15 ps and the production runs consisted out of a 30 ps MD simulation, while the simulation for the hydronium was run for 45 ps, as $\Delta F_{H_3O^+}$ is used for every pK_a calculation.

7 Acknowledgment

The authors would like to thank The Fund for Scientific Research Flanders (FWO) and the Research Board of Ghent University (BOF14/DOC_V/343) for their financial support. Funding was also received from the European Union's Horizon 2020 research and innovation programme [consolidator ERC grant agreement no. 647755–DYNPOR (2015–2020)]. Computational and experimental resources and services used in this work were provided by Ghent University. Special thanks to Louis Vanduyfhuys for the help with YAFF and Toon Verstraelen for the interesting discussions. Support is acknowledged by the Interuniversity Attraction Poles Programme (P7/05) initiated by the Belgian Science Policy Office.

A Appendix

A.1 Collective variables for restrained MD

A well-chosen reaction coordinate is crucial to obtain an accurate free energy profile, of which two are commonly used to study proton transfer reactions (vide infra).[53] These coordinates are preferably chosen to be one-dimensional, since this greatly reduces the amount of simulations that need to be performed. The first commonly used coordinate is a continuous function that estimates the coordination number, n_c, of the number of hydrogens within a cutoff distance r_0 of the donor oxygen atom:[65]

$$n_c = \sum_i \frac{1 - \left(\frac{r(\mathrm{O}-\mathrm{H}_i)}{r_0}\right)^n}{1 - \left(\frac{r(\mathrm{O}-\mathrm{H}_i)}{r_0}\right)^m} \tag{16}$$

where i runs over all hydrogen atoms of the solvent and the acid proton and n and m are constants (in this work respectively 8 and 16).

The argument of the summation assumes a value close to unity when the bond distance $r(\mathrm{O}-\mathrm{H}_i)$ is smaller than r_0 and switches to zero when $r(\mathrm{O}-\mathrm{H}_i) \gg r_0$. For a phenol molecule with r_0 set to 1.2 Å, for instance, only the acid proton will contribute significantly. The coordination number will then be close to one, meaning one proton is bonded to the donor oxygen. For the phenol anion, each term in the summation will have a value close to zero. Therefore, by construction, this function gives an estimate of the amount of protons bonded to the donor oxygen.

This coordinate has the advantage that the acid proton is not specified. As in real water, all protons are treated equally. When simulating low values of n_c, however, the acid proton is transferred to the solution and the coordinate is no longer able to exactly determine the position of the proton. Indeed, a low value of n_c can be achieved not only by one (relatively) close proton that is still partially bonded to the donor oxygen, but also by several small-distance hydrogen bonds.

The second commonly used coordinate includes the distance between the donor oxygen $\mathrm{O_{AH}}$ and the acid proton H, as well as the distance between the acid proton H and the oxygen of the accepting water molecule O_w:[53]

$$\Delta r = r(\mathrm{O_{AH}} - \mathrm{H}) - r(\mathrm{O}_w - \mathrm{H}) \tag{17}$$

in which Δr is the difference between the distance of the proton to the donating oxygen atom (of the acid AH) and the distance of the proton to the accepting water oxygen. If we assume the equilibrium $\mathrm{O_{AH}} - \mathrm{H}$ distance to be approximately equal to 1.0 Å and the equilibrium $\mathrm{O}_w - \mathrm{H}$ distance to be equal to 1.8 Å (a hydrogen bond), this coordinate will vary from −0.8 Å (phenol) through 0.0 Å (when the proton is equidistant between both oxygen atoms) to 0.8 Å (when the phenol is completely deprotonated). The downside of this coordinate is that, when the proton is transferred to the accepting water molecule, another proton of this molecule might jump to a second water molecule and so forth. In this case, the coordinate again is not able to follow the proton once the phenol is deprotonated.

A.2 Restraining potentials and corrections factors for the insertion/deletion scheme

Computationally, the various values of η are sampled through a dummy atom. At $\eta = 0$, this dummy atom is a proton, while at $\eta = 1$, it is essentially a non-interacting particle. This allows us to calculate the energies and forces for AH and A^- as shown in Figure 4. One can thus interpret that for the intermediate values of η, this dummy is a partial proton with a partial positive charge. During a MD trajectory, the position of the dummy atom is governed by its interactions. At lower values of η this is substantial, keeping it apart from other atoms. However, when η approaches unity, the interactions will decrease and vanish ($\eta = 1$), allowing it to move close to other atoms.

When calculating the vertical energy gap, this could lead to extremely large energy values due to bad configurations (e.g., when the dummy's position coincides with another atom). Therefore, restraining potentials are used to keep the dummy atom at positions which would be accessible for a proton. Apart from solving the sampling problem restraining potentials may also be very useful to study strong acids. Without a restraining potential, the proton would quickly diffuse in the solvent medium and it would be very hard to sample the protonated state. Therefore, we opt to use restraining potentials to ensure statistically efficient sampling for all values of η. This procedure applied here was introduced by Costanzo et al.[56].

The extra restraining potentials are chosen to be harmonic:

$$V_r = \sum_{\text{bonds}} \frac{k_r}{2}(r - r_{\text{eq}})^2 + \sum_{\text{angles}} \frac{k_\theta}{2}(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{k_\phi}{2}(\phi - \phi_{\text{eq}})^2 \tag{18}$$

where k_r, k_θ and k_ϕ are force constants. The equilibrium values (denoted by the subscript eq) are determined from the time average of the MD runs. The force constants need to be chosen large enough to keep the dummy in place against thermal fluctuations, but not too large in order to avoid a large bias on the free energies. They are chosen such that fluctuations in bond lengths for instance of the dummy atom are comparable to the fluctuations in simulations with an actual proton. For each phenol derivative, the following parameters were chosen (in a.u.): $r_{\text{eq}} = 1.89\,\text{a}_0$, $k_r = 0.1\,\text{Ha}/\text{a}_0{}^2$, $\theta_{\text{eq}} = 1.94\,\text{rad}$, $k_\theta = 0.1\,\text{Ha}/\text{rad}^2$, $\phi_{\text{eq}} = \pi\,\text{rad}$, $k_\phi = 0.01\,\text{Ha}/\text{rad}^2$; herein a_0 stands for bohr and Ha for hartree, the atomic units for distance and energy respectively. The dihedral angle serves to keep the dummy atom in the plane of the benzene ring. For the pure water simulation, three distance restraining potentials were used to keep the dummy atom and the two other protons bound to one oxygen atom, effectively sampling a hydronium ion: $r_{\text{eq}} = 1.89\,\text{a}_0$, $k_r = 0.1\text{Ha}/\text{a}_0{}^2$.

The use of restraining potentials has an important implication on the free energies. Instead of the deprotonation reaction (1), where the proton is completely removed (and thus presumed to be in the gas phase), we are instead sampling the discharge of this proton into a non-interacting dummy:

$$\text{AH}_{\text{aq}} \rightarrow A d^-_{\text{aq}} \tag{19}$$

This dummy atom is indeed non-interacting, but is still restrained by the above defined potentials. Therefore, the thermodynamic integral (12) will only take into account the discharge of the proton into the neutral dummy atom, but not the (mainly entropic) contribution of releasing this dummy atom into the gas phase. The reader is referred to the appendix of the article by Costanzo et al.[56] for an in-depth discussion of the derivation of the free energy correction factors. In addition, the difference in zero-point energy (ZPE) of the proton bound the acid AH and bound to the hydronium ion can also be taken into account.

With inclusion of these correction factors, the final expression for the pK_a is given by the difference of two thermodynamic integrals (12), one for the acid AH and one for the hydronium ion H_3O^+, one correction factor and a difference of two zero-point energy contributions:

$$\begin{aligned} k_B T \ln 10\ \text{pK}_a =& \Delta F_{\text{AH}} - \Delta F_{\text{H}_3\text{O}^+} + k_B T \ln(c^\circ \Lambda^3_{\text{H}^+}) \\ & - \Delta_{zp} E_{\text{H(A)}} + \Delta_{zp} E_{\text{H(OH}_2)^+} \end{aligned} \tag{20}$$

with $c^\circ = 1\text{mol/L}$ the unit molar concentration and Λ_{H^+} the thermal wavelength of a proton. Herein, ΔF_{AH} and $\Delta F_{\text{H}_3\text{O}^+}$ can be determined using the thermodynamic integral (12) discussed above. $k_B T \ln(c^\circ \Lambda^3_{\text{H}^+})$ is a constant taking into account the entropy change associated with the release of the dummy atom attached to AH into the gas phase. This term equals to -3.2 pK_a units.

To accurately determine the ZPE contributions of the proton, path integral dynamics within the solvent environment would be required. This is beyond the scope of this paper, which is why the ZPE contributions are estimated as follows:

$$\begin{aligned} \Delta_{zp}E_{\mathrm{H(A)}} &= E_{zp}(\mathrm{AH}) - E_{zp}(\mathrm{A}^-) \\ \Delta_{zp}E_{\mathrm{H(OH_2)^+}} &= E_{zp}(\mathrm{H_3O^+}) - E_{zp}(\mathrm{H_2O}) \end{aligned} \tag{21}$$

The ZPE energy corrections (E_{zp}) in the above equations are easily determined through a static calculation. In this work, this was performed in the Gaussian software package at the IEF-PCM-M06-2X/6-311g(d,p) level of theory.[91, 92] As the Δ_{zp} values obtained this way are only an estimate of the real ZPE of the proton, we will report pK_a's calculated with and without ZPE corrections, denoted $\mathrm{pK}_{a,i/d}$ and $\mathrm{pK}_{a,i/d}{}^*$ respectively (in which the subscript i/d denotes the insertion/deletion scheme).

B Keywords

molecular dynamics, steered MD, dyes, pK_a, free energy methods

C TOC

An acidity constant (pK_a) is a very important quantity in solution chemistry. The goal of this contribution is to predict the pK_a of large solvated dye molecules. Hereto, two advanced MD simulations are compared and evaluated for a set of phenol derivatives: restrained MD (see Figure 12) and the insertion/deletion scheme. After evaluation, the insertion/deletion scheme is applied to the solvated dye molecule Bromothymol Blue, with excellent agreement to experiment.

References

[1] D Staneva, R Betcheva, and J M Chovelon. Fluorescent benzo de anthracen-7-one pH-sensor in aqueous solution and immobilized on viscose fabrics. *J. Photoch. Photobio. A*, 183(1-2):159–164, 2006.

[2] D Staneva, R Betcheva, and J M Chovelon. Optical sensor for aliphatic amines based on the simultaneous colorimetric and fluorescence responses of smart textile. *J. Appl. Polym. Sci.*, 106(3):1950–1956, 2007.

[3] L. Van der Schueren and K. De Clerck. The Use of pH-indicator Dyes for pH-sensitive Textile Materials. *Text. Res. J.*, 80(7):590–603, October 2009.

[4] Lien Van der Schueren, Tybo Mollet, Özgür Ceylan, and Karen De Clerck. The development of polyamide 6.6 nanofibres with a pH-sensitive function by electrospinning. *Eur. Polym. J.*, 46(12):2229–2239, December 2010.

[5] L Van der Schueren and K De Clerck. Textile materials with a pH-sensitive function. *Int. J. Cloth. Sci. Tech.*, 23(4):269–274274, 2011.

[6] Guowei Li, Jia Xiao, and Wenqin Zhang. A novel dual colorimetric fiber based on two acid–base indicators. *Dyes Pigm.*, 92(3):1091–1099, March 2012.

[7] Lien Van der Schueren and Karen De Clerck. Coloration and application of pH-sensitive dyes on textile materials. *Color. Technol.*, 128(2):82–90, April 2012.

[8] Lien Van der Schueren, Karen De Clerck, Giovanna Brancatelli, Giuseppe Rosace, Els Van Damme, and Winnok De Vos. Novel cellulose and polyamide halochromic textile sensors based on the encapsulation of Methyl Red into a sol–gel matrix. *Sens. Actuators, B*, 162(1):27–34, February 2012.

[9] Lien Van der Schueren, Karen Hemelsoet, Veronique Van Speybroeck, and Karen De Clerck. The influence of a polyamide matrix on the halochromic behaviour of the pH-sensitive azo dye Nitrazine Yellow. *Dyes Pigm.*, 94(3):443–451, September 2012.

[10] Lien Van Der Schueren, Thierry De Meyer, Iline Steyaert, Özgür Ceylan, Karen Hemelsoet, Veronique Van Speybroeck, and Karen De Clerck. Polycaprolactone and polycaprolactone/chitosan nanofibres functionalised with the pH-sensitive dye Nitrazine Yellow. *Carbohydr. Polym.*, 91(1):284–293, 2013.

[11] Iline Steyaert, Gertjan Vancoillie, Richard Hoogenboom, and Karen De Clerck. Dye immobilization in halochromic nanofibers through blend electrospinning of a dye-containing copolymer and polyamide-6. *Polym. Chem.*, 6(14):2685–2694, 2015.

[12] Lars Alexander Schneider, Andreas Korber, Stephan Grabbe, and Joachim Dissemond. Influence of pH on wound-healing: A new perspective for wound-therapy? *Arch. Dermatol. Res.*, 298(9):413–420, 2007.

[13] Junming Ho and ML Coote. p K a Calculation of Some Biologically Important Carbon Acids-An Assessment of Contemporary Theoretical Procedures. *J. Chem. Theory Comput.*, 5:295–306, 2009.

[14] Junming Ho and Michelle L. Coote. A universal approach for continuum solvent pKa calculations: Are we there yet? *Theor. Chem. Acc.*, 125(1-2):3–21, 2010.

[15] George C. Shields and Paul G. Seybold. *Computational Approaches for the Prediction of pKa Values*. CRC Press, 2013.

[16] C J Cramer and D G Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99(8):2161–2200, 1999.

[17] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.*, 105(8):2999–3094, 2005.

[18] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B*, 113(18):6378–6396, 2009.

[19] Josefredo R. Pliego and José M. Riveros. Theoretical Calculation of p K a Using the Cluster−Continuum Model. *J. Phys. Chem. A*, 106:7434–7439, 2002.

[20] M D Liptak and G C Shields. Accurate pK(a) calculations for carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *J. Am. Chem. Soc.*, 123(30):7314–7319, August 2001.

[21] Matthew D. Liptak, Kevin C. Gross, Paul G. Seybold, Steven Feldgus, and George C. Shields. Absolute pKa determinations for substituted phenols. *J. Am. Chem. Soc.*, 124(22):6421–6427, 2002.

[22] Josefredo R. Pliego. Thermodynamic cycles and the calculation of pKa. *Chem. Phys. Lett.*, 367(1-2):145–149, 2003.

[23] Juan Crugeiras, Ana Ríos, and Howard Maskill. DFT and AIM study of the protonation of nitrous acid and the p K a of nitrous acidium ion. *J. Phys. Chem. A*, 115(44):12357–12363, 2011.

[24] Catherine C R Sutton, George V. Franks, and Gabriel Da Silva. First principles pK a calculations on carboxylic acids using the SMD solvation model: Effect of thermodynamic cycle, model chemistry, and explicit solvent molecules. *J. Phys. Chem. B*, 116(39):11999–12006, 2012.

[25] Jafar Zanganeh, Mohammednoor Altarawneh, Ibrahim Saraireh, Samira Namazi, and Javad Zanganeh. Theoretical study on thermochemical parameters and pKa values for fluorinated isomers of toluene. *Comput. Theor. Chem.*, 1011:21–29, 2013.

[26] Satesh Gangarapu, Antonius T M Marcelis, and Han Zuilhof. Accurate pKa calculation of the conjugate acids of alkanolamines, alkaloids and nucleotide bases by quantum chemical methods. *Chem. Phys. Chem.*, 14(5):990–995, 2013.

[27] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J. Phys. Chem. B*, 110(32):16066–16081, 2006.

[28] Vyacheslav S. Bryantsev, Mamadou S. Diallo, and William A. Goddard. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *J. Phys. Chem. B*, 112(32):9709–9719, 2008.

[29] Farhad Khalili, Amr Henni, and Allan L L East. Entropy contributions in pKa computation: Application to alkanolamines and piperazines. *J. Mol. Struct. THEOCHEM*, 916(1-3):1–9, 2009.

[30] Aida Mariana Rebollar-Zepeda and Annia Galano. First principles calculations of p K a values of amines in aqueous solution: Application to neurotransmitters. *Int. J. Quantum Chem.*, 112(21):3449–3460, November 2012.

[31] Shuming Zhang. A reliable and efficient first principles-based method for predicting pK a values. III. Adding explicit water molecules: Can the theoretical slope be reproduced and pK a values predicted more accurately? *J. Comput. Chem.*, 33(5):517–526, 2012.

[32] Joo-Eun Jee, Aleix Comas-Vives, Chiara Dinoi, Gregori Ujaque, Rudi van Eldik, Agustí Lledós, and Rinaldo Poli. Nature of Cp*MoO2+ in water and intramolecular proton-transfer mechanism by stopped-flow kinetics and density functional theory calculations. *Inorg. Chem.*, 46(10):4103–13, 2007.

[33] Junming Ho and Michelle L. Coote. First-principles prediction of acidities in the gas and solution phase. *WIREs Comput Mol Sci*, 1(5):649–660, September 2011.

[34] R Car and M Parrinello. Unified Approach for Molecular-Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, 55(22):2471–2474, 1985.

[35] D Marx and J Hutter. *Ab initio molecular dynamics - Basic Theory and Advanced Methods*. Cambridge, 2009.

[36] Veronique Van Speybroeck and Robert J. Meier. A recent development in computational chemistry: chemical reactions from first principles molecular dynamics simulations. *Chem. Soc. Rev.*, 32(3):151–157, 2003.

[37] S. L. C. Moors, B. Brigou, D. Hertsen, B. Pinter, P. Geerlings, V. Van Speybroeck, S. Catak, and F. De Proft. Influence of Solvation and Dynamics on the Mechanism and Kinetics of Nucleophilic Aromatic Substitution Reactions in Liquid Ammonia. *J. Org. Chem.*, 81(4):1635–1644, 2016.

[38] Thierry De Meyer, Karen Hemelsoet, Veronique Van Speybroeck, and Karen De Clerck. Substituent effects on absorption spectra of pH indicators: An experimental and computational study of sulfonphthaleine dyes. *Dyes Pigm.*, 102:241–250, 2014.

[39] Thierry De Meyer, Iline Steyaert, Karen Hemelsoet, Richard Hoogenboom, Veronique Van Speybroeck, and Karen De Clerck. Halochromic properties of sulfonphthaleine dyes in a textile environment: The influence of substituents. *Dyes Pigm.*, 124:249–257, 2016.

[40] D Marx, M E Tuckerman, J Hutter, and M Parrinello. The nature of the hydrated excess proton in water. *Nature*, 397(6720):601–604, 1998.

[41] P L Geissler, C Dellago, D Chandler, J Hutter, and M Parrinello. Autoionization in liquid water. *Science*, 291(5511):2121–2124, 2001.

[42] Mark E Tuckerman, Dominik Marx, and Michele Parrinello. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, 417(6892):925–929, 2002.

[43] D Asthagiri, L R Pratt, and J D Kress. Ab initio molecular dynamics and quasichemical study of H+(aq). *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6704–6708, 2005.

[44] H Grubmüller, B Heymann, and P Tavan. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.

[45] S Izrailev, S Stepaniants, M Balsera, Y Oono, and K Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72(4):1568–1581, 1997.

[46] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72(4):1541–1555, 1997.

[47] M Balsera, S Stepaniants, S Izrailev, Y Oono, and K Schulten. Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.*, 73(3):1281–1287, 1997.

[48] Alicia Claudia Lorenzo and Ernesto Ra??l Caffarena. Elastic properties, Young's modulus determination and structural stability of the tropocollagen molecule: A computational study by steered molecular dynamics. *J. Biomech.*, 38(7):1527–1533, 2005.

[49] George Patargias, Hugo Martay, and Wolfgang B Fischer. Reconstructing potentials of mean force from short steered molecular dynamics simulations of Vpu from HIV-1. *J. Biomol. Struct. Dyn.*, 27(1):1–12, 2009.

[50] Vignir Ísberg, Thomas Balle, Tommy Sander, Flemming Steen Jørgensen, and David E. Gloriam. G protein- and agonist-bound serotonin 5-HT2A receptor model activated by steered molecular dynamics simulations. *J. Chem. Inf. Model.*, 51(2):315–325, 2011.

[51] Giacomo Fiorin, Michael L. Klein, and Jérôme Hénin. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.*, 111(22-23):3345–3362, 2013.

[52] Michiel Sprik. Computation of the pk of liquid water using coordination constraints. *Chem. Phys.*, 258(23):139–150, 2000.

[53] Murat Kılıç and Bernd Ensing. Acidity constants of lumiflavin from first principles molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, 16(35):18993–19000, 2014.

[54] Marialore Sulpizi and Michiel Sprik. Acidity constants from vertical energy gaps: density functional theory based molecular dynamics implementation. *Phys. Chem. Chem. Phys.*, 10(34):5238–5249, 2008.

[55] Jun Cheng, Marialore Sulpizi, and Michiel Sprik. Redox potentials and pKa for benzoquinone from density functional theory based molecular dynamics. *J. Chem. Phys.*, 131(15):154504, 2009.

[56] Francesca Costanzo, Marialore Sulpizi, Raffaele Guido Della Valle, and Michiel Sprik. The oxidation of tyrosine and tryptophan studied by a molecular dynamics normal hydrogen electrode. *J. Chem. Phys.*, 134(24):244508, 2011.

[57] Jun Cheng, Chengden Liu, J VandeVondele, Marialore Sulpizi, and Michiel Sprik. Redox potentials and acidity constants from density functional theory based molecular dynamics. *Acc. Chem. Res.*, 47:3522–3529, 2014.

[58] R. A. Marcus and N. Sutin. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta*, 811(3):265–322, 1985.

[59] Gregory King and Arieh Warshel. Investigation of the free energy functions for electron transfer reactions. *J. Chem. Phys.*, 8682, 1990.

[60] Marialore Sulpizi and Michiel Sprik. Acidity constants from DFT-based molecular dynamics simulations. *J. Phys - Condens. Mat.*, 22(28):284116, 2010.

[61] Xiandong Liu, Xiancai Lu, Michiel Sprik, Jun Cheng, Evert Jan Meijer, and Rucheng Wang. Acidity of edge surface sites of montmorillonite and kaolinite. *Geochim. Cosmochim. Acta*, 117:180–190, 2013.

[62] Thierry De Meyer, Karen Hemelsoet, Lien Van Der Schueren, Ewald Pauwels, Karen De Clerck, and Veronique Van Speybroeck. Investigating the halochromic properties of azo dyes in an aqueous environment by using a combined experimental and theoretical approach. *Chem. Eur. J.*, 18(26):8120–8129, 2012.

[63] E.a. Carter, Giovanni Ciccotti, James T. Hynes, and Raymond Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156(5):472–477, 1989.

[64] W. K. Den Otter and W. J. Briels. The calculation of free-energy differences by constrained molecular-dynamics simulations. *J. Chem. Phys.*, 109(11):4139–4146, 1998.

[65] Michiel Sprik. Coordination numbers as reaction coordinates in constrained molecular dynamics. *Faraday Discuss.*, 110(0):437–445, 1998.

[66] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180(10):1961–1972, 2009.

[67] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185(2):604–613, 2014.

[68] Phillip L. Geissler, Christoph Dellago, and David Chandler. Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B*, 103(18):3706–3710, 1999.

[69] Gareth A. Tribello, Michele Ceriotti, and Michele Parrinello. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *P. Natl. Acad. Sci. U.S.A.*, 109(14):5196–5201, 2012.

[70] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.*, 67(1):159–184, 2016.

[71] Yasuhiro Matsunaga, Yasuaki Komuro, Chigusa Kobayashi, Jaewoon Jung, Takaharu Mori, and Yuji Sugita. Dimensionality of collective variables for describing conformational changes of a multi-domain protein. *J. Phys. Chem. Lett.*, 7(8):1446–1451, 2016.

[72] M. Abramowitz and I. Setegun. *A handbook of mathematical functions.* National Bureau of Standards, 1965.

[73] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipies Third Edition.* Cambridge University Press, 2007.

[74] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103–128, 2005.

[75] G Lippert, J Hutter, and M Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.*, 92(3):477–487, 1997.

[76] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, 1988.

[77] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, 1988.

[78] S Grimme, J Antony, S Ehrlich, and H Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15):154104, 2010.

[79] S Goedecker, M Teter, and J Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54(3):1703–1710, 1996.

[80] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *Chem. Phys.*, 127(11):114105, 2007.

[81] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[82] G J Martyna, M L Klein, and M Tuckerman. Nosé-Hoover Chains: The Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.*, 97(4):2635–2643, 1992.

[83] Joost VandeVondele, Fawzi Mohamed, Matthias Krack, Jürg Hutter, Michiel Sprik, and Michele Parrinello. The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.*, 122(1):14515, 2005.

[84] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and Testing of a General Amber Force Field. *J. Comput. Chem.*, 56531:1157–1174, 2004.

[85] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.*, 25(2):247–260, 2006.

[86] Leandro Martínez, R. Andrade, E. G. Birgin, and José Mario Martínez. Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.*, 31(16):2967–2970, 2010.

[87] T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande, and S. M. J. Rogge. Yaff, yet another force field, 2016.

[88] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[89] H Flyvbjerg and H G Petersen. Error-Estimates on Averages of Correlated Data. *J. Chem. Phys.*, 91(1):461–466, 1989.

[90] Legault, C. Y. CYLview, 1.0b, 2009. Université de Sherbrooke (http://www.cylview.org).

[91] M J Frisch, G W Trucks, H B Schlegel, G E Scuseria, M A Robb, J R Cheeseman, G Scalmani, V Barone, B Mennucci, G A Petersson, H Nakatsuji, M Caricato, X Li, H P Hratchian, A F Izmaylov, J Bloino, G Zheng, J L Sonnenberg, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, T Vreven, J A Montgomery Jr., J E Peralta, F Ogliaro, M Bearpark, J J Heyd, E Brothers, K N Kudin, V N Staroverov, R Kobayashi, J Normand, K Raghavachari, A Rendell, J C Burant, S S Iyengar, J Tomasi, M Cossi, N Rega, J M Millam, M Klene, J E Knox, J B Cross, V Bakken, C Adamo, J Jaramillo, R Gomperts, R E Stratmann, O Yazyev, A J Austin, R Cammi, C Pomelli, J W Ochterski, R L Martin, K Morokuma, V G Zakrzewski, G A Voth, P Salvador, J J Dannenberg, S Dapprich, A D Daniels, Ö Farkas, J B Foresman, J V Ortiz, J Cioslowski, and D J Fox. *Gaussian 09, Revision D.* Gaussian, Inc., Wallingford CT, 2009.

[92] Y Zhao and D G Truhlar. 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 function. *Theor. Chem. Acc.*, 120(1-3):215–241, 2008.

[93] E Serjeant and B Dempsey. *Ionisation constants of organic acids in aqueous solution.* Pergamon Press, 1979.

[94] J. Drahonovsky and Z. Vacek. Dissoziationskonstanten und Austauscherchromatographie Chlorierter Phenole. *Czechoslov. Chem. Commun*, 36:3431–3440, 1971.

[95] G. Kortüm, W. Vogel, and K. Andrussow. *Dissociation constants of organic acids in aqueous solution.* Butterworth, 1961.

[96] P J Pearce and R J J Simkins. Acid strengths of some substituted picric acids. *Can. J. Chem.*, 46(2):241–248, 1968.

10.1002/cphc.201600734

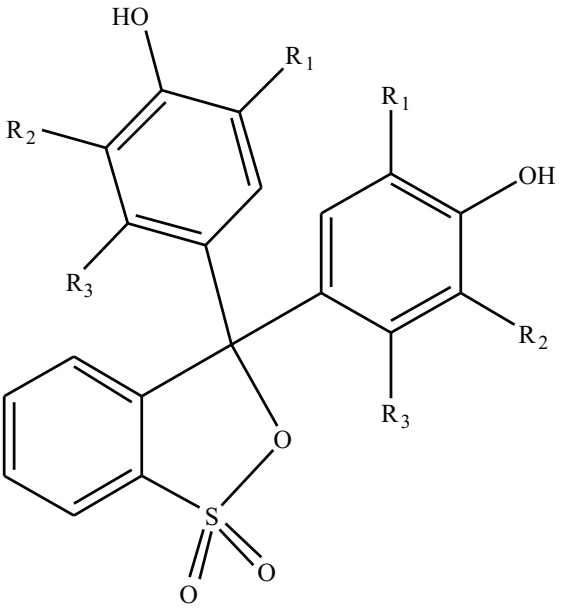

Figure 1: General structure of the sulfonphthaleine dye class; for Bromothymol Blue: R_1=*i*-Pr, R_2=Br, R_3=CH_3.

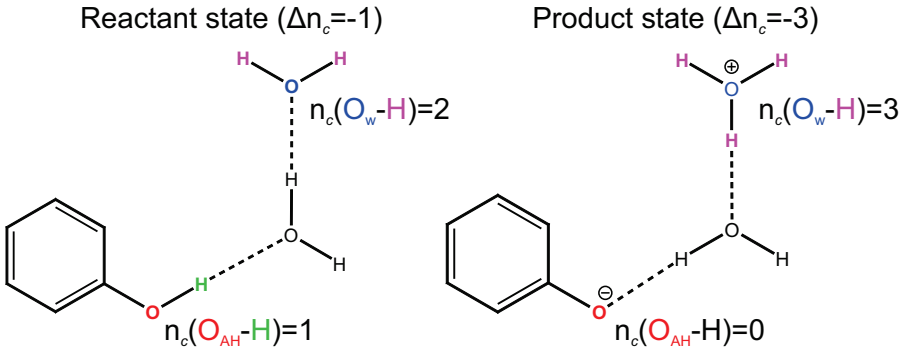

Figure 2: In the reactant state (protonated phenol), the coordination number of O_{AH} is close to one, while the coordination number of O_w (from a water molecule in the second solvation layer) is close to two. Δn_c is thus equal to -1. In the product state, the phenol is deprotonated and the water molecule has become a hydronium ion, meaning Δn_c is equal to -3. Forcing Δn_c from -1 to -3 by restraining potentials, will thus cause O_{AH} to deprotonate. This proton will be transferred through a proton wire, with the water molecule from the first solvation layer as a "bridge", to the accepting oxygen atom O_w.

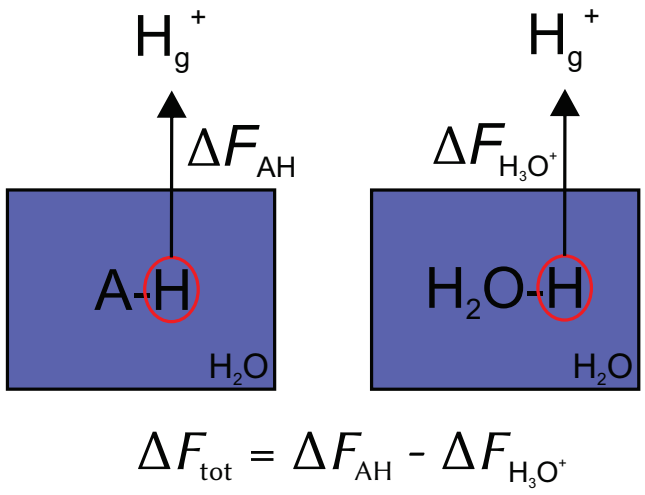

Figure 3: Graphical representation of the insertion/deletion scheme.

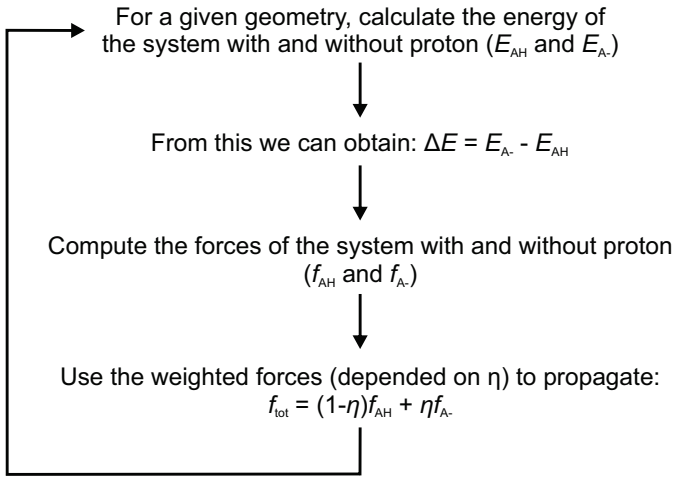

Figure 4: This workflow for the insertion/deletion scheme allows us to effectively sample different values of η while also obtaining $\langle \Delta E \rangle_\eta$.

10.1002/cphc.201600734

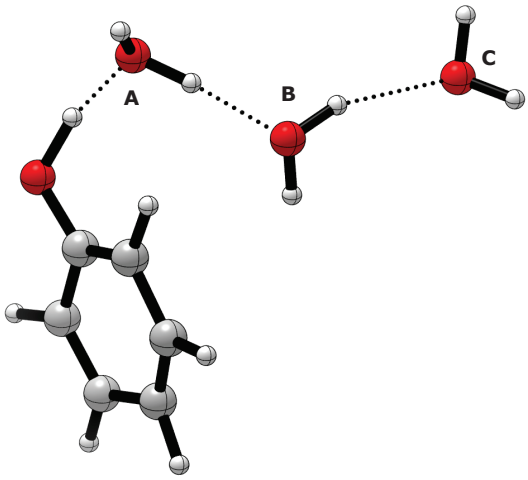

Figure 5: Labeling of water molecules **A**, **B** and **C** in respectively the first, second and third solvation layer.

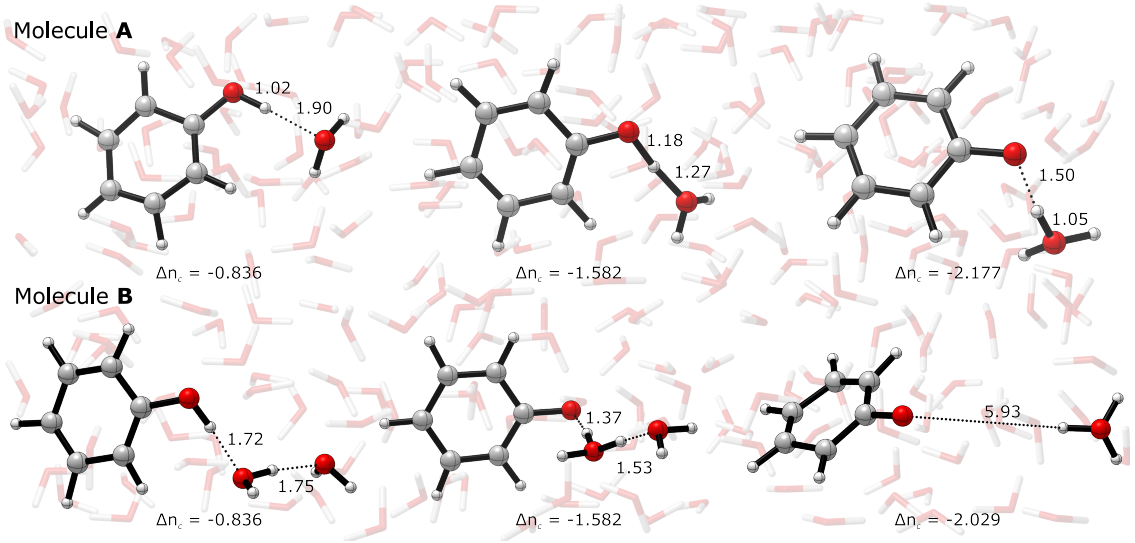

Figure 6: Snapshots during the restrained MD simulations for phenol with respectively **A** and **B** as hydronium. From left to right these snapshots show the start, middle and end of the reaction; distances are given in Å.

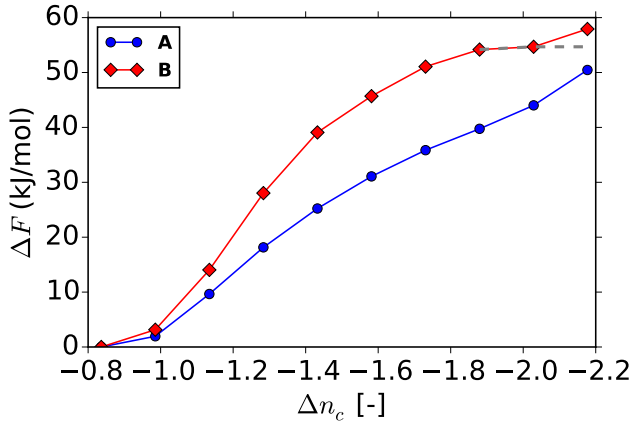

Figure 7: Free energy differences (ΔF) as a function of the difference in coordination number Δn_c, calculated from restrained MD simulations with **A** and **B** chosen as hydronium.

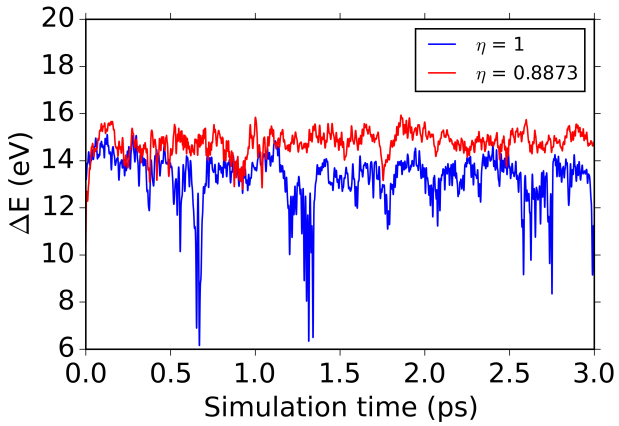

Figure 8: ΔE for a simulation of phenol for $\eta = 1$ and 0.8873 during an initial 3 ps simulation. For $\eta = 0.8873$, much lower oscillations are observed, which will lead to better convergence due to better sampling.

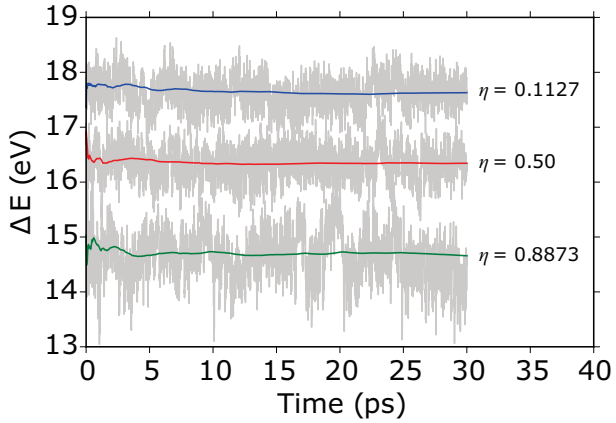

Figure 9: The running averages of ΔE for all values of η are shown for phenol. In grey, the actual values are also given. The latter show large fluctuations, explaining why relatively long MD runs are necessary to obtain statistically converged results.

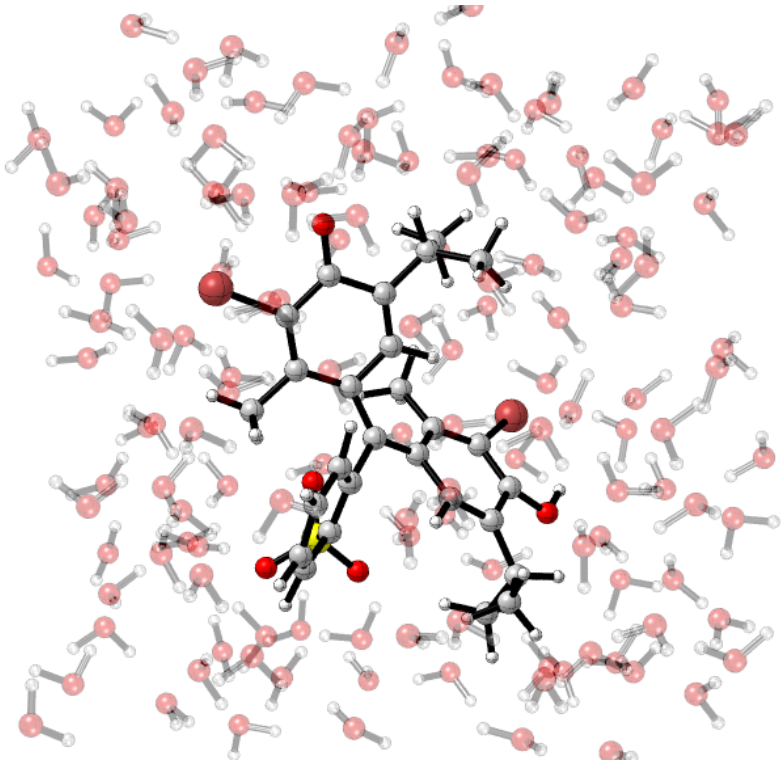

Figure 10: The sulfonphthaleine dye Bromothymol Blue is solvated by 137 water molecules in a 16.863 $Å^3$ cubic box.

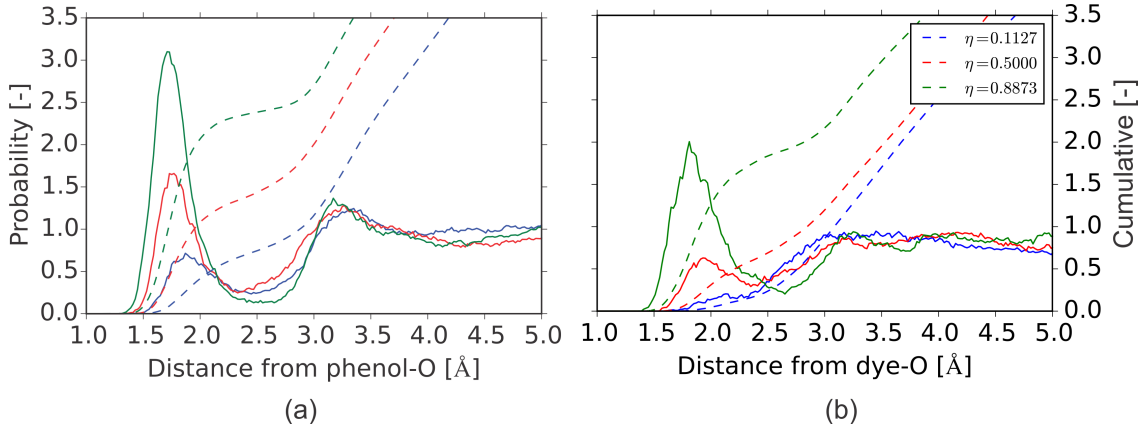

Figure 11: Radial distribution function (RDF) and cumulative RDF for phenol (a) and Bromothymol Blue (b) between the donating oxygen atom and all solvent hydrogens. The curves are calculated for all three values of η (0.1127, 0.5 and 0.8873).

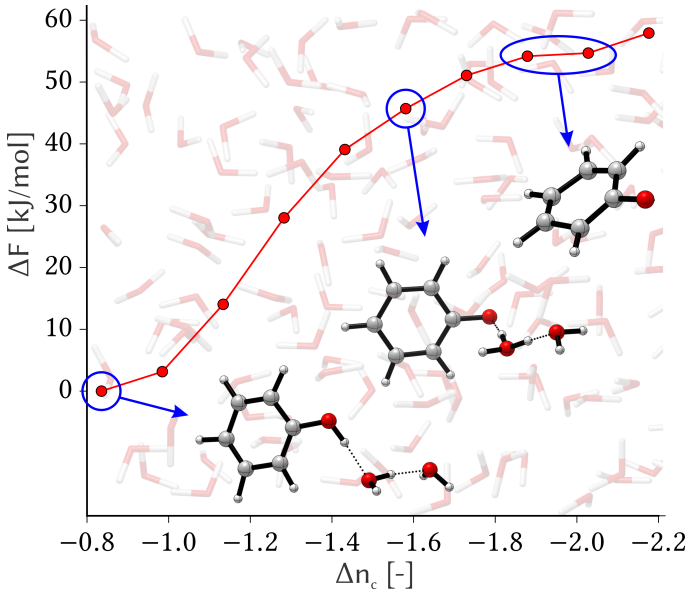

Figure 12: The free energy profile is shown as a function of the reaction coordinate, namely a difference in coordination number Δn_c. This coordinate is successful in sampling the entire deprotonation reaction.

Table 1: List of molecules considered in this work and their experimental acidity constants

Compound		Experimental pK_a
1	phenol	10.0[93]
2	2-chlorophenol	8.5[94]
3	3-chlorophenol	9.1[93]
4	4-chlorophenol	9.4[93]
5	2-bromophenol	8.5[95]
6	2,6-dibromophenol	6.7[93]
7	2,4-dibromophenol	7.8[93]
8	2-methylphenol	10.3[96]
9	3-methylphenol	10.1[96]
10	4-hydroxybenzoic acid	9.3[93]
11	Bromothymol Blue	7.4[38]

Table 2: pK_a's calculated with restrained MD ($pK_{a,restr}$) and the insertion/deletion method with and without zero-point energy corrections ($pK_{a,i/d}$ and $pK_{a,i/d}{}^*$ respectively) compared to experimental values (pK_a, exp); standard deviations are also given

Compound		$pK_{a,restr}$	$pK_{a,i/d}$	$pK_{a,i/d}{}^*$	$pK_{a,exp}$
1	phenol	9.7 ± 0.7	9.0 ± 0.9	9.1 ± 0.9	10.0
2	2-chlorophenol	7.5 ± 0.9	7.3 ± 1.0	8.1 ± 1.0	8.5
3	3-chlorophenol	8.5 ± 1.3	7.9 ± 1.1	8.3 ± 1.1	9.1
4	4-chlorophenol	9.3 ± 0.8	9.0 ± 1.0	9.2 ± 1.0	9.4
5	2-bromophenol	8.2 ± 0.7	7.8 ± 1.1	8.1 ± 1.1	8.5
6	2,6-dibromophenol	6.6 ± 0.4	7.2 ± 1.0	7.6 ± 1.0	6.7
7	2,4-dibromophenol	7.1 ± 0.8	8.2 ± 1.0	8.8 ± 1.0	7.8
8	2-methylphenol	9.8 ± 1.1	9.4 ± 1.1	9.7 ± 1.1	10.3
9	3-methylphenol	10.4 ± 1.1	8.9 ± 1.0	8.9 ± 1.0	10.1
10	4-hydroxybenzoic acid	9.0 ± 1.0	8.8 ± 1.0	9.0 ± 1.0	9.3