Acidity Constant (pK_a) Calculation of Large Solvated Dye Molecules: Evaluation of Two Advanced Molecular Dynamics Methods

Thierry De Meyer,[a, b] Bernd Ensing,[c] Sven M. J. Rogge,[a] Karen De Clerck,[b] Evert Jan Meijer,*[c] and Veronique Van Speybroeck*[a]

pH-Sensitive dyes are increasingly applied on polymer substrates for the creation of novel sensor materials. Recently, these dye molecules were modified to form a covalent bond with the polymer host. This had 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 the future intelligent design of sensor materials. Herein, we show that advanced molecular dynamics (MD) methods have reached the level at which the pK_a values of large solvated dye molecules can be predicted with high accuracy. Two MD methods were used in this work: steered or restrained MD and the insertion/deletion scheme. Both were first calibrated on a set of phenol derivatives and afterwards applied to the dye molecule bromothymol blue. Excellent agreement with experimental values was obtained, which opens perspectives for using these methods for designing dye molecules.

1. 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, for which 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 [Eq. (1)]:

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

[a] T. De Meyer, S. M. J. Rogge, Prof. V. Van Speybroeck
     Center for Molecular Modeling, Ghent University
     Technologiepark 903, 9052 Zwijnaarde (Belgium)
     E-mail: veronique.vanspeybroeck@ugent.be

[b] T. De Meyer, Prof. K. De Clerck
     Department of Textiles, Ghent University
     Technologiepark 907, 9052 Zwijnaarde (Belgium)

[c] Prof. B. Ensing, Prof. E. J. Meijer
     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

     Supporting Information and the ORCID identification number(s) for the
     author(s) of this article can be found under http://dx.doi.org/10.1002/
     cphc.201600734.

of which the pK_a is given by [Eq. (2)]:

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

in which ΔG^o_a is the standard dissociation free energy, k_B is the Boltzmann constant, and T is the absolute temperature. This free-energy difference between the protonated and deprotonated state is the central quantity that needs to be calculated.

Our specific interest in calculating pK_a values arises from the development of novel sensor materials: halochromic (pH-sensitive) dyes can be incorporated into polymeric structures to give pH-sensitive polymers that can be used in wound bandages, protective clothing, and so on.[1–10] In recent developments, these dye molecules were modified to include reactive or polymerizable groups.[11] This resulted in a material for which the dye was covalently bound to the host material, which greatly improved the leaching properties. This modification, however, changed 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 behavior. Ideally, dye molecules could be designed to yield halochromic behavior in the desired pH region. This requires techniques that are not only able to explain experimentally observed pH-sensitive properties but that are also 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.[13–15] Thanks to developments in solvent reaction field methods,[16–18] for which the solute is placed inside a cavity in a dielectric continuum, the pK_a values of molecules in solution have been successfully calculated by 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.[20–26] It was previously shown that adding explicit solvent molecules inside the cavity can improve the accuracy of calculated pK_a values, although it is challenging to determine the necessary amount of water molecules.[27–33]

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 by treating both the compound and the 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 (DFTMD) simulations.[34,35] The advantage of DFTMD is the uniform treatment of the entire system, which allows not only electronic polarization of the solvent and solute but also explicit interactions such as hydrogen bonds. Herein, we use such techniques, as 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 visualized in Figure 1.[38,39] The next step in de-
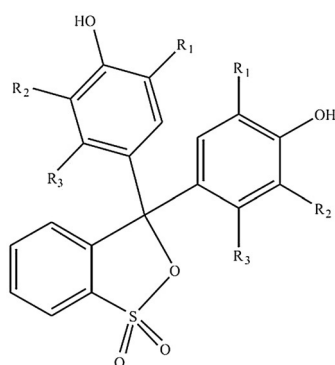
Figure 1. General structure of the sulfonphthaleine dye class; for bromothymol blue: $R_1 = i$Pr, $R_2 = $Br, and $R_3 = CH_3$.

signing 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 used two advanced molecular dynamics (MD) methods that fully captured the effects of the molecular environment, which were 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 bore substituents similar to those found in sulfonphthaleine dyes (Table 1) and were previously studied by static approaches.[21] Afterwards, the most suited procedure was applied to bromothymol blue,

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

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

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 value of 7.4.[38]

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

2. Advanced MD Techniques for pK_a Determination

Proton-transfer reactions have been the subject of various studies.[40–43] The difficulty in studying these reactions with MD methods is that proton transfer is a rare event. Advanced MD methods are, therefore, required to simulate these processes, as the timescale of a MD run is typically of the order of a few picoseconds. 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[44–51] and was pioneered by Sprik and co-workers specifically for pK_a calculations.[52] The idea behind this method is to integrate the force needed to "drag" the system from the protonated state 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 2.1 and Appendix A.1.

The group of Sprik developed another interesting approach to calculate pK_a values.[54–57] This method concerns the reversible elimination of protons and electrons, which allows redox potentials and acidity constants to be calculated on the basis of the principles of the Marcus theory of electron transfer.[58,59] The pK_a calculation itself is based on an insertion/deletion scheme, for which Reaction (1) is split into two half-reactions and the free-energy change can be determined on the basis of thermodynamic integration. This method has been successfully applied to several aqueous systems,[60] and Cheng et al. concluded that an accuracy of 1–2 pK_a units is feasible.[57,61] This method is discussed in Section 2.2 and Appendix A.2.

It was previously shown that performing MD simulations can greatly improve the accuracy of calculated absorption wavelengths of dye molecules (in combination with time-dependent DFT) compared to static calculations.[62] As will be shown here, MD simulations are also effective for calculating the pK_a values

of large solvated systems. In Section 2, the theory of restrained MD and the insertion/deletion method will be discussed. In Section 3, both methods are applied to phenol derivatives, and the results will be discussed. Afterwards, the solvated dye bromothymol blue will be studied.

2.1. 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 (ΔF) between the reactant and product states is equal to the work required to shift the system along the chosen coordinate; it can be calculated by integrating the force (f) along the coordinate [Eq. (3)]:[53,65]

$$k_B T \ln 10 \, pK_a = \Delta F = -\int_{q_0}^{q_1} \langle f \rangle_q \, dq \qquad (3)$$

in which we set $F(q_0) = 0$. The parameter q needs to be a well-chosen reaction coordinate, in which q_0 and q_1 correspond to the reactant and product states, 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, the Hess law ensures that the resulting pK_a 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.[68–71] 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 acidic 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 atom) to the aqueous solution. We will assume that this occurs in the form of a hydronium ion, but this will be discussed further on. As was observed in simulations with the abovementioned collective variables,[53] the acidic 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 for which the anion and hydronium ion are as far away from each other as possible, which would allow 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 acidic proton will increase. Therefore, we propose to use a difference in coordination number as the reaction coordinate [Eq. (4)]:

$$\Delta n_c = n_c(O_{AH}) - n_c(O_w)$$
$$= \sum_i \frac{1 - \left(\frac{r(O_{AH}-H_i)}{r_0}\right)^n}{1 - \left(\frac{r(O_{AH}-H_i)}{r_0}\right)^m} - \sum_i \frac{1 - \left(\frac{r(O_w-H_i)}{r_0}\right)^n}{1 - \left(\frac{r(O_w-H_i)}{r_0}\right)^m} \qquad (4)$$

in which n_c is the coordination number and r_0 is the cutoff distance. The choice of the accepting water molecule will be discussed in Section 3.1. For a water molecule in the second solvation layer, this coordinate is illustrated in Figure 2.
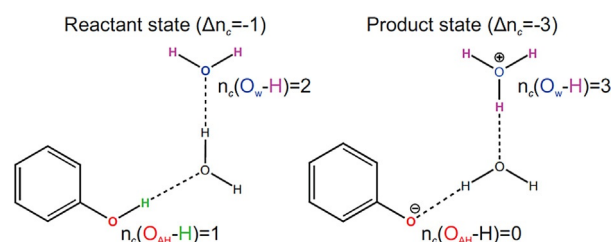
Figure 2. In the reactant state (protonated phenol), the coordination number of O_{AH} is close to 1, whereas the coordination number of O_w (from a water molecule in the second solvation layer) is close to 2. Δn_c is thus equal to -1. In the product state, the phenol is deprotonated and the water molecule has become a hydronium ion; consequently, 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.

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 if 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 in which both the donating and accepting oxygen atoms tend to have a proton in their proximity. The acidic proton will therefore remain between both oxygen atoms, which would put some strain on the proton wire. The smoothing parameters n and m are chosen to be 8 and 16, respectively, which results in a relatively steep switching function. It is also important to mention that for 4-hydroxybenzoic acid, a quadratic wall is used to keep the coordination numbers of both oxygen atoms of the carboxylate group low; otherwise, the acidic 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.

2.2. Insertion/Deletion

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

Total Reaction (1) is split up into two half-reactions [Eqs. (5) and (6)]:

$$AH_{(aq)} \rightarrow A^-_{(aq)} + H^+_{(g)} \qquad (5)$$

$$H^+_{(aq)} \rightarrow H^+_{(g)} \qquad (6)$$

Herein, an intermediate step in the gas phase 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
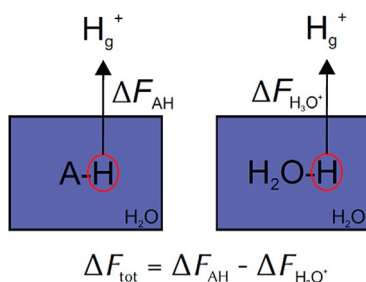

Figure 3. Graphical representation of the insertion/deletion scheme.

the two simulations, the total free-energy change can be obtained. The derivation followed here is written down for the acid AH, but it is identical for a proton in solution (e.g. in the form of a hydronium ion, 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 [Eq. (7)]:

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

For $\eta = 0$, the system is in the protonated state (i.e. AH), whereas for $\eta = 1$ the deprotonated state (A^-) is retrieved. The intermediate values correspond to a mixture of AH and A^-; this implies 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 [Eq. (8)]:

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

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 [Eq. (9)]:

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

in which N is the number of particles, Λ is the thermal wave-length, and $\beta = (k_B T)^{-1}$.

The free-energy derivative can then be written as an ensemble average [Eq. (10)]:

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

Employing Equation (7), the derivative of the internal energy can be written as [Eq. (11)]:

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

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 illustrated in the workflow given in Figure 4. Each value of η defines a dif-
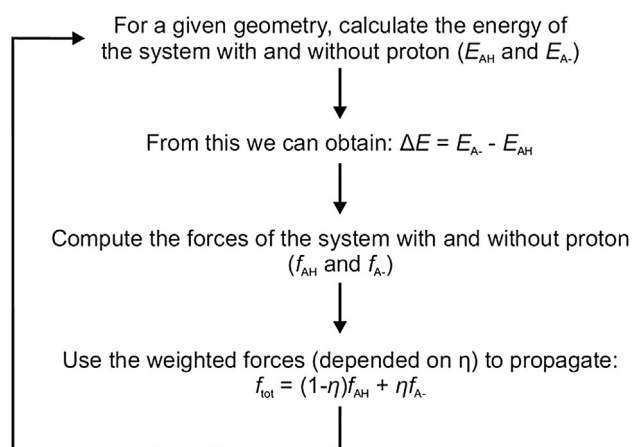

Figure 4. This workflow for the insertion/deletion scheme allows us to effectively sample different values of η while also obtaining $\langle\Delta E\rangle_\eta$.

ferent mixing and, therefore, also requires a different MD simulation. As a result, one obtains accumulative averaging energy gaps as a function of η (see below).

Finally, the free-energy difference [Eq. (8)] can be written as [Eq. (12)]:

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

Consequently, 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 [Eq. (12)] can be obtained as [Eq. (13)]:

$$\Delta F = \frac{1}{2}\left(\langle \Delta E \rangle_0 + \langle \Delta E \rangle_1\right) = \frac{1}{2}\left(\Delta E_{AH} + \Delta E_{A^-}\right) \qquad (13)$$

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

If the proton is inserted/removed, this will, in general, give rise to a substantial reorganization of the surrounding solvent molecules. Nonlinear effects are therefore expected and the numerical integration of Equation (12) may require a better quadrature by using a three-point integration formula (evaluating in $\eta = 0$, 0.5 and 1) [Eq. (14)]:

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

Instead of the three-point trapezium quadrature (Simpson's rule), a Gauss–Legendre quadrature can also be used [Eq. (15)]:

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

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 3.2). The practical implementation and final pK_a expression are discussed in Appendix A.2.

3. Results and Discussion

Table 1 lists the molecules under study with their experimental pK_a values. For each molecule, the acidity constants will be calculated and rationalized. The pK_a values of the different phenol derivatives can be easily understood from the electron-donating/withdrawing effect of their substituent(s).[38] The halogen atoms in compounds 2–7 cause the pK_a to drop with respect to unfunctionalized phenol (1). The electron-withdrawing effect of these substituents results in decreased charge density in the aromatic system; consequently, a more acidic environment is needed to stabilize the acid proton. A methyl group has a small electron-donating effect, which translates into slightly higher pK_a values for compounds 8 and 9. Compound 10 has a slightly lower pK_a value than 1, which indicates that the carboxylate group, even though it is already negatively charged, still allows for some stabilization of the extra negative charge upon deprotonation of 10. For chlorine-substituted phenols 2–4, a clear trend in 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.

3.1. Restrained MD

In this work, a difference in coordination number (Δn_c) is used as the reaction coordinate, as explained in the previous section. The value of Δn_c is calculated between the donating oxygen atom of the phenol (the acid AH) and the oxygen atom of the accepting water molecule. The most important choice to be made is the selection of the water molecule to which the proton is transferred, which hence becomes the hydronium ion. Phenol is taken as a reference system, and three different water molecules are taken as candidates to become the hydronium ion. The first choice is clear: 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
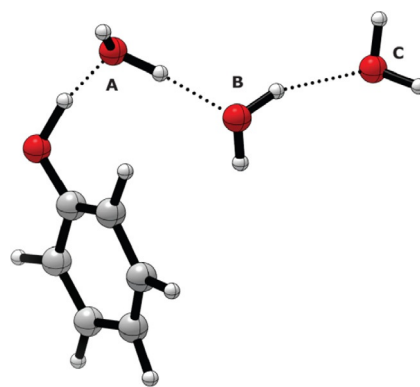

Figure 5. Labeling of water molecules **A**, **B**, and **C** in the first, second, and third solvation layers, respectively.

a molecule in the second solvation layer, which makes this molecule our second choice (B). Following the hydrogen-bonding network, a third water molecule (C) is also chosen. The larger the distance between the phenol anion and the hydronium ion, the more stable each ion can be in their own solvation layer. From the three selected molecules, however, simulations for which we select molecule C are unstable. The acid proton does not remain between phenol and the chosen water molecule and escapes to the surrounding solvent. Consequently, the collective variable as constructed is not able to maintain the proton wire. For water molecules A and B, however, the simulations are stable.

Two different simulations were then performed: with Δn_c defined as the difference between the coordination number of the phenol oxygen atom (OPh) and the oxygen atom of water molecule A (OA) in one simulation and with Δn_c between OPh and the oxygen atom of water molecule B (OB) in the other. The value of Δn_c was varies between −0.836 and −2.177. These values correspond to protonated and deprotonated phenol and are 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
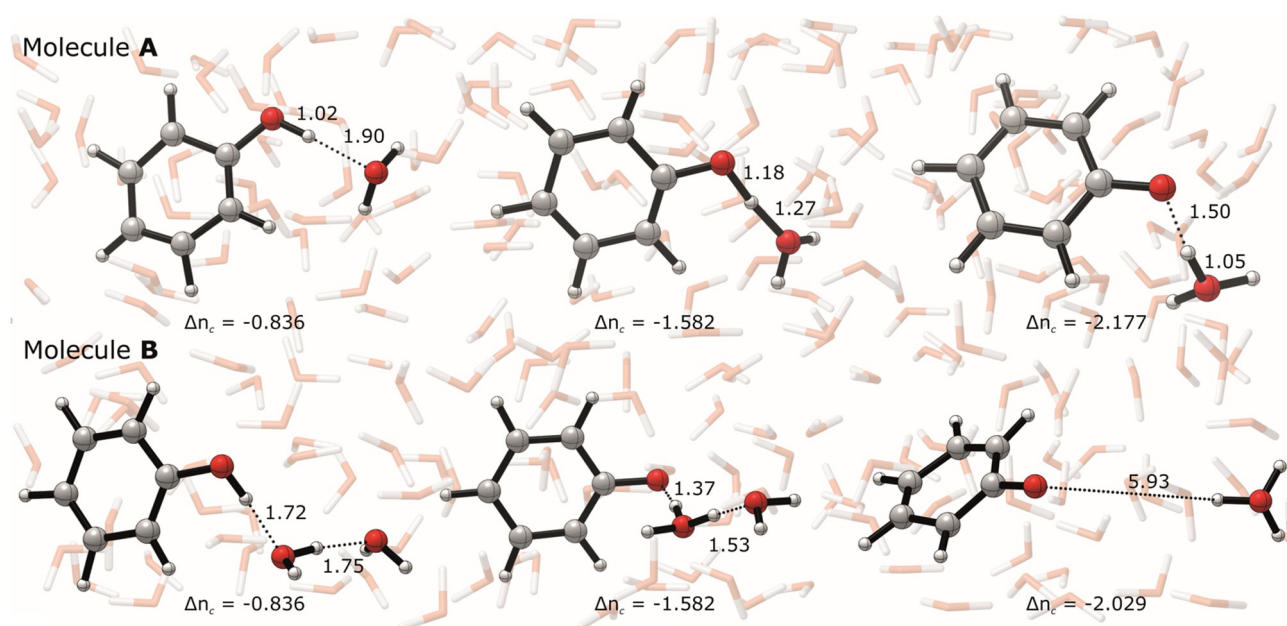
Figure 6. Snapshots during the restrained MD simulations for phenol with **A** and **B** as the hydronium ion. From left to right these snapshots show the start, middle, and end of the reaction; distances are given in Å.

oxygen atom repelling neighboring protons, whereas 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 neighboring water molecule. For **B**, the proton first hops to the closest water molecule. All molecules are rather close to each other, which corresponds to a proton wire. In the final snapshot, the reaction is complete and the hydronium ion and phenol anion have moved away from each other. This is not observed in situation **A**, which can be understood upon 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
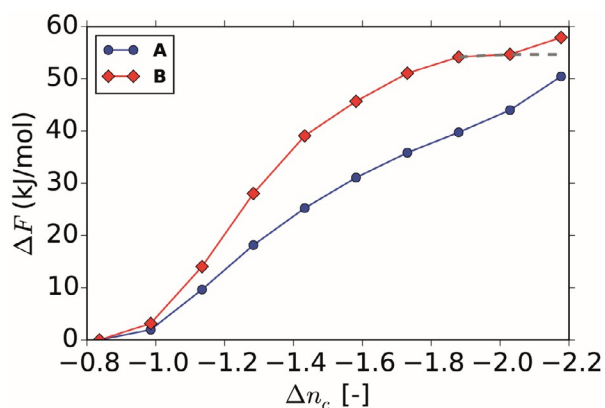
Figure 7. Free-energy differences (ΔF) as a function of the difference in coordination number Δn_c, calculated from restrained MD simulations with **A** and **B** chosen as the hydronium ion.

a point at which the "stable" product is formed. Ions have the tendency to diffuse away from each other in aqueous solution, which is not possible with Δn_c as constructed for **A**; this explains why no plateau is formed. For **B**, however, a plateau is reached between the eighth and ninth points, as the free energy barely changes, which corresponds with a stable state. The absence of a clear product valley in the free-energy profile, whereas the reaction has clearly finished (as can be seen from Figure 6), may be explained as follows. Phenol is an alkalic compound and thus the most stable state will always correspond to that for which 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 favored. In the current setup of the system with a solvated box of 64 water molecules, it would be impossible to simulate such effects. For further simulations with the use of the restrained MD method, we use the simulation for which **B** is selected as the accepting water molecule and choose the energy corresponding to the plateau to deduce the free-energy difference. In all simulations, this state corresponds to a slightly lower coordination number for the accepting oxygen atom (OB). This allows the hydronium ion to reside in Zundel or Eigen ionic structures, which indeed have been shown to be lower in free energy.[40]

Using the full free-energy profile for the simulation in which **A** is chosen as the hydronium ion, ΔF is 50.5 ± 4.1 kJ mol^{-1}, which corresponds to a pK_a value of 8.9 ± 0.7. For the simulation in which **B** is the hydronium, ΔF is 54.7 ± 4.0 kJ mol^{-1} upon taking into account points 1 to 9 (and omitting the final point), which corresponds to a pK_a value 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 ion is there

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

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

a clearly observable plateau, which makes the end point for the integration interval nonarbitrary. Therefore, for 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; energy values can be found in the Supporting Information.

Overall, the results obtained with the restrained MD simulations reproduce quite reliably the experimentally observed trend in the acidity constants. Some simulations, however, show the same instability as that observed with molecule **C** in the case of phenol. More specifically, for molecules **3**, **4**, and **7** the proton escapes to the solvent upon simulating at $\Delta n_c = -1.73$. Molecule **6** proves more problematic: the simulation is also unstable at $\Delta n_c = -1.43$ and -1.88. As mentioned in Section 2, the proton wire is only stable if both oxygen atoms (between which Δn_c is calculated) still want the acidic proton in their proximities. These compounds have lower pK_a values than phenol; consequently, they will have a lower inclination to have the proton close by. These points correspond to the situation in which the phenol is completely deprotonated but for which the chosen water molecule cannot yet become a hydronium ion (due to the chosen value of Δn_c). For molecule **6**, both bromine substituents also give steric hindrance for the formation of a proton wire, which makes it extra difficult to sample the reaction, as performed here. Therefore, the above-mentioned points are omitted upon calculating the pK_a values, and it is stressed that only stable simulations are used for the final results. Nevertheless, the results obtained with the 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, as 2.5 pK_a units is generally considered as chemically accurate.[14]

Even though the results are in excellent agreement with the experimental values, the use of the difference in coordination number has two downsides. The first is general to all restrained MD simulations: the limited box size. As already mentioned, the phenol anion and hydronium ion are most stable if they can be far away from each other; this allows each to have its own solvation shell. Second, the simulation is unstable at some points of Δn_c, as the proton does not remain in the origi-

nally defined proton wire but rather diffuses away. These shortcomings will be addressed in the following section by employing the insertion/deletion method.

3.2. Insertion/Deletion

In the insertion/deletion scheme, 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, which makes it unnecessary to sample a proton wire, as was the case with restrained MD. Therefore, the insertion/deletion scheme can address both previously found shortcomings.

The two quadratures mentioned in Section 2.2 to estimate the thermodynamic integral [Eq. (12)] are evaluated here. For the trapezium rule [Eq. (14)], the average energy difference $\langle \Delta E \rangle$ between the protonated and deprotonated systems is evaluated during simulations of η equal to 0, 0.5, and 1. For $\eta = 1$, the dummy atom is a noninteracting particle. Given that 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 noninteracting particle and equal to a proton. This might lead to energetically very unfavorable conformations, which can be seen from Figure 8: for the simulation at $\eta = 1$, strong negative peaks can
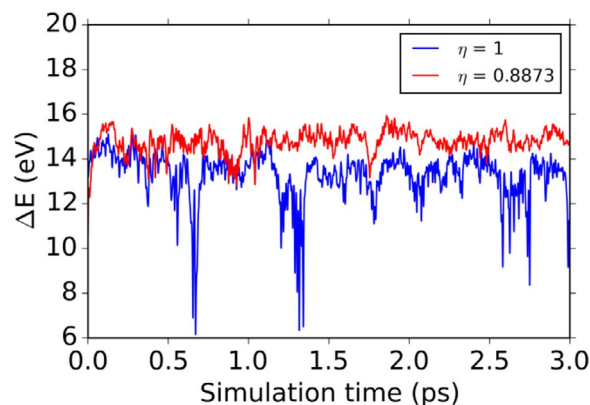
Figure 8. ΔE for a simulation of phenol for $\eta = 1$ and 0.8873 during an initial 3 ps simulation. For $\eta = 0.8873$, much lower oscillations are observed, which will lead to better convergence due to better sampling.

be seen throughout the simulation. This is easy to understand from a chemical point of view: these peaks correspond to geometries in which the dummy atom is very close to the proton of a water molecule that forms a hydrogen bond with the negatively charged oxygen atom.

The Gauss–Legendre quadrature [Eq. (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. Chem-
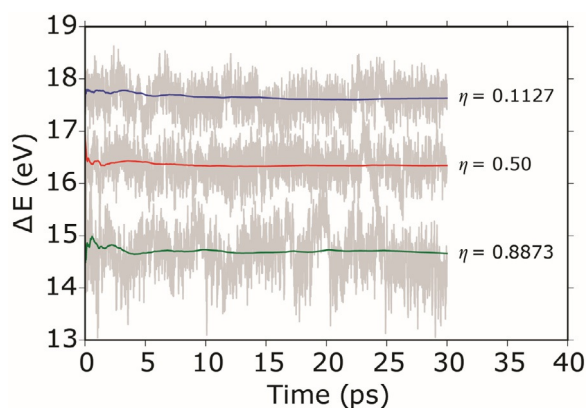
Figure 9. The running averages of ΔE for all values of η are shown for phenol. In gray, the actual values are also given. The latter show large fluctuations, which explains why relatively long MD runs are necessary to obtain statistically converged results.

Figure 10. The sulfonphthaleine dye bromothymol blue is solvated by 137 water molecules in a 16.863 Å^3 cubic box.

ically, this implies 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 $pK_{a,i/d}$ values are in good agreement with the experimental ones. The root-mean-square difference between experiment and theory is 0.9 pK_a units. Including zero-point energy (ZPE) corrections can have a large effect on the final pK_a: its effect can be negligible or amount 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 Supporting Information. The standard deviation on the pK_a values is on average 1.0. Summarizing, 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 is also evaluated and is discussed in the Supporting Information. Ho and Coote recommend use 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 Supporting Information, as such methods are not feasible and are beyond the scope of this study. In the previous part, the restrained MD method was shown to provide accurate results relative to experiment, but the insertion/deletion scheme provided the most stable simulations, as certain shortcomings found with restrained MD could be circumvented. Therefore,
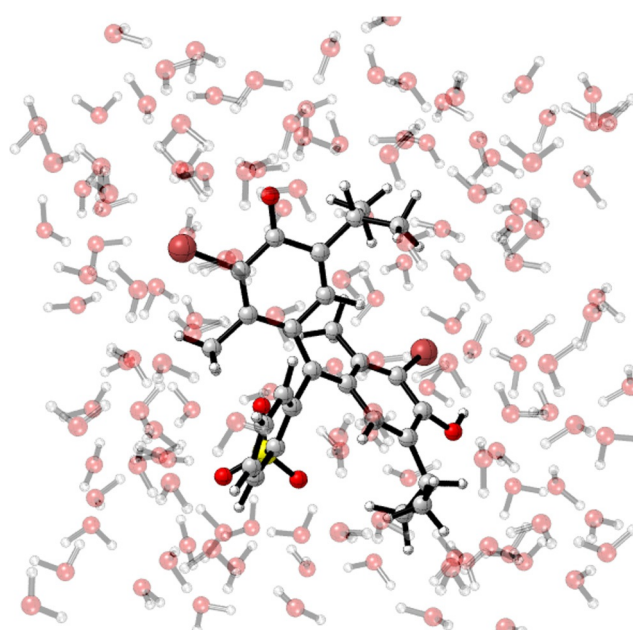
preference was given to this method for application to the dye molecule bromothymol blue.

A snapshot during the MD simulation of the solvated bromothymol blue system is shown in Figure 10. The system consists of 473 atoms in total, which is very large and explains the slight adaptation in computational parameters (see Section 5). The pK_a value 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 are also calculated, but in this case, they have 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 (RDFs) 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 that indicates the probability of finding a solvent proton at a certain distance from the donating oxygen atom. 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 a cumulative RDF (CRDF), 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 are made. With $\eta = 0.1127$, the RDF shows a relatively small peak at a distance of around 1.87 Å. This peak defines the position of the first solvation layer around the phenol oxygen atom, 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 atom is still bonded to the dummy atom (which is almost a full proton) and, therefore, has only one free
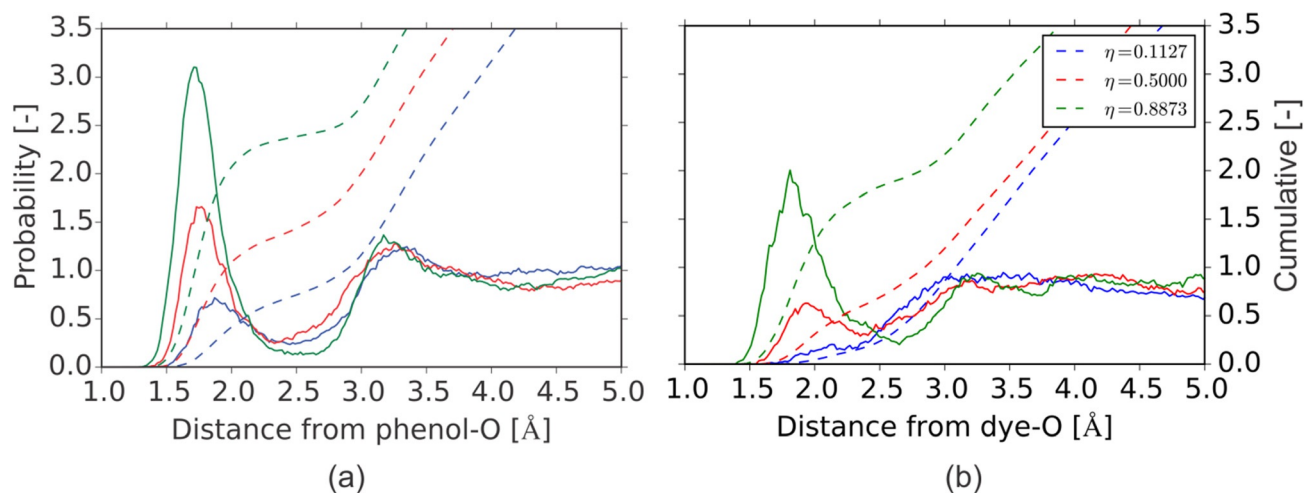
Figure 11. Radial distribution function (RDF) and cumulative RDF for a) phenol and b) bromothymol blue between the donating oxygen atom and all solvent hydrogen atoms. The curves are calculated for all three values of η (0.1127, 0.5, and 0.8873).

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 noninteracting (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 that 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 heights of the peaks. This can be attributed to steric 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 a distance of 1.93 Å, which contains on average 0.6 water molecules. If 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 Å with on average 1.84 water molecules. Note that this is still less than that observed for phenol (2.3 water molecules), which again shows the effect of steric 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 the pK_a values of modified dye molecules to be predicted and the effect of different, more complex environments to be studied.

4. 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 in which the sensor is active. In this contribution, we showed that advanced molecular dynamics (MD) were capable of calculating pK_a values for large solvated dye molecules.

Two MD-based methods were discussed and first evaluated on a set of phenol derivatives. The first method was restrained MD, for which a novel reaction coordinate was proposed. The commonly used coordinates, namely, a coordination number or a difference in distance, showed difficulties in maintaining control of the entire reaction. Therefore, a different coordinate
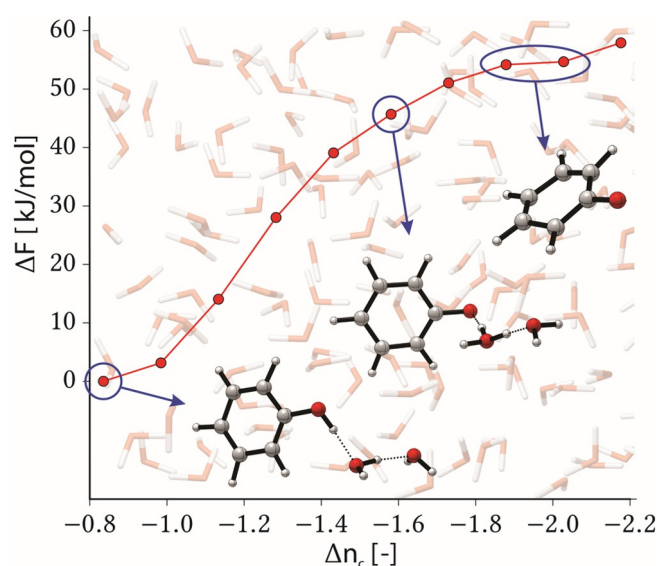
Figure 12. The free-energy profile is shown as a function of the reaction coordinate, namely, a difference in coordination number, Δn_c. This coordinate is successful in sampling the entire deprotonation reaction.

was utilized, namely, a difference in coordination numbers, Δn_c. This coordinate allowed us to sample the full deprotonation reaction (Figure 12), 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, which makes this approach very promising.

The second method was the insertion/deletion scheme, as originally proposed by the group of Sprik. This procedure has the advantage that deprotonation of the phenol and protonation of a water molecule (to a hydronium ion) are sampled in different simulation boxes. No issues with stability of the simulation arose, and the phenol and hydronium ion could 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 relative 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, which thus confirmed the insertion/deletion scheme to be a reliable method for calculating the pK_a values 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 the pK_a values of modified dyes or dyes in more complex environments in the future.

Computational Details

All simulations in this work were performed by using the CP2K/Quickstep package.[74] Electronic structures were calculated with density functional theory, which was implemented on the basis of a hybrid Gaussian plane wave (GPW) approach.[75] The BLYP functional was used for the exchange correlation,[76,77] whereas Grimme D3 dispersion corrections were used to improve van der Waals interactions.[78] GTH pseudopotentials were 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, wavefunction optimization tolerance was set to 0.3×10^{-7} Hartree. To enlarge the MD time step, we replaced the hydrogens atoms by deuterium atoms, which allowed energy conservation with a 0.75 fs time step to be maintained. Temperature was initially brought to the desired 320 K with a CSVR thermostat, after which the temperature was maintained at 320 K by using a Nosé-Hoover chain thermostat with four beads.[81,82] The higher temperature was chosen to avoid the glassy (over structured) behavior of BLYP liquid water found at lower temperatures.[83]

The models used in this work consisted of 3D periodically repeated cubic cells with a lattice parameter of 12.462 Å. This corresponded to the volume of 64 water molecules at experimental density, which was used for the pure water simulations. For the solvated

phenols, the same system size was used. This allowed us to rely on error compensation upon calculating the total free-energy change for the insertion/deletion scheme. From the solvated phenol systems, an amount of water molecules was removed to maintain atmospheric pressure. This amount was estimated by performing several force-field simulations at fixed volume with a different amount of water molecules and averaging the pressure by 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, whereas 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 interacted with the solvent.

Bromothymol blue required a larger solvent box, namely, a cubic cell with a 16.863 Å lattice size, which corresponded 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, which was 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] visualization of the MD simulations was done with VMD.[88] Block averaging methods were used to estimate standard deviations on the obtained forces and $\langle \Delta E \rangle$ values.[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 of a 30 ps MD simulation, whereas the simulation for the hydronium was run for 45 ps, as $\Delta F_{H_3O^+}$ was used for every pK_a calculation.

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 (see below).[53] These coordinates are preferably chosen to be one dimensional, as 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 hydrogen atoms within a cutoff distance r_0 of the donor oxygen atom [Eq. (16)]:[65]

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

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

The argument of the summation assumes a value close to unity if the bond distance $r(O-H_i)$ is smaller than r_0 and switches to zero if

$r(\text{O–H}) \gg r_0$. For a phenol molecule with r_0 set to 1.2 Å, for instance, only the acidic proton will contribute significantly. The coordination number will then be close to one; consequently, one proton is bonded to the donor oxygen atom. 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 atom.

This coordinate has the advantage that the acid proton is not specified. As in real water, all protons are treated equally. Upon simulating low values of n_c, however, the acidic 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 atom but also by several small-distance hydrogen bonds.

The second commonly used coordinate includes the distance between the donor oxygen atom O_{AH} and the acidic proton H, as well as the distance between the acid proton H and the oxygen atom of the accepting water molecule O_w [Eq. (17)]:[53]

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

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 atom. If we assume the equilibrium O_{AH}–H distance to be approximately equal to 1.0 Å and the equilibrium O_w–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 (if the proton is equidistant between both oxygen atoms) to 0.8 Å (if the phenol is completely deprotonated). The downside of this coordinate is that if 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, whereas at $\eta = 1$, it is essentially a noninteracting 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, which keeps it apart from other atoms. However, if η approaches unity, the interactions will decrease and vanish ($\eta = 1$), which would allow it to move close to other atoms. Upon calculating the vertical energy gap, this could lead to extremely large energy values due to bad configurations (e.g. if the dummy's position coincides with another atom). Therefore, restraining potentials are used to keep the dummy atom at positions that 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 into 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 [Eq. (18)]:

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

in which 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 they must not be too large 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 au): $r_{\text{eq}} = 1.89\,a_0$, $k_r = 0.1\,\text{Ha}\,a_0^{-2}$, $\theta_{\text{eq}} = 1.94\,\text{rad}$, $k_\theta = 0.1\,\text{Ha}\,\text{rad}^{-2}$, $\phi_{\text{eq}} = \pi\,\text{rad}$, $k_\phi = 0.01\,\text{Ha}\,\text{rad}^{-2}$; 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: $r_{\text{eq}} = 1.89\,a_0$, $k_r = 0.1\,\text{Ha}\,a_0^{-2}$.

The use of restraining potentials has an important implication on the free energies. Instead of the deprotonation reaction [Eq. (1)], for which 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 noninteracting dummy [Eq. (19)]:

$$\text{AH}_{(aq)} \rightarrow \text{Ad}^-_{(aq)} \tag{19}$$

This dummy atom is indeed noninteracting, but it is still restrained by the above-defined potentials. Therefore, the thermodynamic integral [Eq. (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 to the acid AH and bound to the hydronium ion can also be taken into account.

With the inclusion of these correction factors, the final expression for the pK_a is given by the difference in two thermodynamic integrals [Eq. (12)], one for the acid AH and one for the hydronium ion H_3O^+, one correction factor, and the difference in two zero-point energy contributions [Eq. (20)]:

$$k_B T \ln 10\, pK_a = \Delta F_{AH} - \Delta F_{H_3O^+} + k_B T \ln(c^o \Lambda_{H^+}^3)$$
$$- \Delta_{zp} E_{H(A)} + \Delta_{zp} E_{H(OH_2)^+} \tag{20}$$

for which $c^o = 1\,\text{mol}\,L^{-1}$ is the unit molar concentration and Λ_{H^+} is the thermal wavelength of a proton. Herein, ΔF_{AH} and $\Delta F_{H_3O^+}$ can be determined by using the thermodynamic integral [Eq. (12)] discussed above. The $k_B T \ln(c^o \Lambda_{H^{+3}})$ term 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 −3.2 pK_a units.

To accurately determine the ZPE contributions of the proton, path integral dynamics within the solvent environment would be re-

quired. This is beyond the scope of this paper, which is why the ZPE contributions are estimated as follows [Eq. (21)]:

$$\Delta_{zp}E_{H(A)} = E_{zp}(AH) - E_{zp}(A^-)$$
$$\Delta_{zp}E_{H(OH_2)^+} = E_{zp}(H_3O^+) - E_{zp}(H_2O)$$

(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 values calculated with and without ZPE corrections, denoted p$K_{a,i/d}$ and p$K_{a,i/d}^*$, respectively (in which the subscript i/d denotes the insertion/deletion scheme).

Acknowledgment

Keywords: acidity · dyes/pigments · free energy methods · molecular dynamics · steered molecular dynamics

[1] D. Staneva, R. Betcheva, J. M. Chovelon, *J. Photochem. Photobiol. A* **2006**, *183*, 159–164.
[2] D. Staneva, R. Betcheva, J. M. Chovelon, *J. Appl. Polym. Sci.* **2007**, *106*, 1950.
[3] L. Van der Schueren, K. De Clerck, *Text. Res. J.* **2010**, *80*, 590.
[4] L. Van der Schueren, T. Mollet, Ö. Ceylan, K. De Clerck, *Eur. Polym. J.* **2010**, *46*, 2229.
[5] L. Van der Schueren, K. De Clerck, *Int. J. Cloth. Sci. Tech.* **2011**, *23*, 269.
[6] G. Li, J. Xiao, W. Zhang, *Dyes Pigm.* **2012**, *92*, 1091.
[7] L. Van der Schueren, K. De Clerck, *Color. Technol.* **2012**, *128*, 82.
[8] L. Van der Schueren, K. De Clerck, G. Brancatelli, G. Rosace, E. Van Damme, W. De Vos, *Sens. Actuators B* **2012**, *162*, 27.
[9] L. Van der Schueren, K. Hemelsoet, V. Van Speybroeck, K. De Clerck, *Dyes Pigm.* **2012**, *94*, 443.
[10] L. Van Der Schueren, T. De Meyer, I. Steyaert, Ö. Ceylan, K. Hemelsoet, V. Van Speybroeck, K. De Clerck, *Carbohydr. Polym.* **2013**, *91*, 284.
[11] I. Steyaert, G. Vancoillie, R. Hoogenboom, K. De Clerck, *Polym. Chem.* **2015**, *6*, 2685.
[12] L. A. Schneider, A. Korber, S. Grabbe, J. Dissemond, *Arch. Dermatol. Res.* **2007**, *298*, 413.
[13] J. Ho, M. L. Coote, *J. Chem. Theory Comput.* **2009**, *5*, 295.
[14] J. Ho, M. L. Coote, *Theor. Chem. Acc.* **2010**, *125*, 3.
[15] G. C. Shields, P. G. Seybold, *Computational Approaches for the Prediction of pK$_a$ Values*, CRC Press, Boca Raton, FL, **2013**.
[16] C. J. Cramer, D. G. Truhlar, *Chem. Rev.* **1999**, *99*, 2161.
[17] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.
[18] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378.
[19] J. R. Pliego, J. M. Riveros, *J. Phys. Chem. A* **2002**, *106*, 7434.
[20] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

[21] M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421.
[22] J. R. Pliegom, *Chem. Phys. Lett.* **2003**, *367*, 145.
[23] J. Crugeiras, A. Ríos, H. Maskill, *J. Phys. Chem. A* **2011**, *115*, 12357.
[24] C. C. R. Sutton, G. V. Franks, G. Da Silva, *J. Phys. Chem. B* **2012**, *116*, 11999.
[25] J. Zanganeh, *Comput. Theor. Chem.* **2013**, *1011*, 21.
[26] S. Gangarapu, A. T. M. Marcelis, H. Zuilhof, *ChemPhysChem* **2013**, *14*, 990.
[27] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2006**, *110*, 16066.
[28] V. S. Bryantsev, M. S. Diallo, W. A. Goddard, *J. Phys. Chem. B* **2008**, *112*, 9709.
[29] F. Khalili, A. Henni, A. L. L. East, *J. Mol. Struct. THEOCHEM* **2009**, *916*, 1.
[30] A. M. Rebollar-Zepeda, A. Galano, *Int. J. Quantum Chem.* **2012**, *112*, 3449.
[31] S. Zhang, *J. Comput. Chem.* **2012**, *33*, 517.
[32] J.-E. Jee, A. Comas-Vives, C. Dinoi, G. Ujaque, R. van Eldik, A. Lledós, R. Poli, *Inorg. Chem.* **2007**, *46*, 4103.
[33] J. Ho, M. L. Coote, *WIREs Comput. Mol. Sci.* **2011**, *1*, 649.
[34] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471.
[35] D. Marx, J. Hutter, *Ab Initio Molecular Dynamics—Basic Theory and Advanced Methods*, Cambridge University Press, New York, **2009**.
[36] V. Van Speybroeck, R. J. Meier, *Chem. Soc. Rev.* **2003**, *32*, 151.
[37] S. L. C. Moors, B. Brigou, D. Hertsen, B. Pinter, P. Geerlings, V. Van Speybroeck, S. Catak, F. De Proft, *J. Org. Chem.* **2016**, *81*, 1635.
[38] T. De Meyer, K. Hemelsoet, V. Van Speybroeck, K. De Clerck, *Dyes Pigm.* **2014**, *102*, 241.
[39] T. De Meyer, I. Steyaert, K. Hemelsoet, R. Hoogenboom, V. Van Speybroeck, K. De Clerck, *Dyes Pigm.* **2016**, *124*, 249.
[40] D. Marx, M. E. Tuckerman, J. Hutter, M. Parrinello, *Nature* **1999**, *397*, 601.
[41] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, M. Parrinello, *Science* **2001**, *291*, 2121.
[42] M. E. Tuckerman, D. Marx, M. Parrinello, *Nature* **2002**, *417*, 925.
[43] D. Asthagiri, L. R. Pratt, J. D. Kress, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6704.
[44] H. Grubmüller, B. Heymann, P. Tavan, *Science* **1996**, *271*, 997.
[45] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten, *Biophys. J.* **1997**, *72*, 1568.
[46] E. Evans, K. Ritchie, *Biophys. J.* **1997**, *72*, 1541.
[47] M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono, K. Schulten, *Biophys. J.* **1997**, *73*, 1281.
[48] A. C. Lorenzo, E. R. Caffarena, *J. Biomech.* **2005**, *38*, 1527.
[49] G. Patargias, H. Martay, W. B. Fischer, *J. Biomol. Struct. Dyn.* **2009**, *27*, 1.
[50] V. Ísberg, T. Balle, T. Sander, F. Steen Jørgensen, D. E. Gloriam, *J. Chem. Inf. Model.* **2011**, *51*, 315.
[51] G. Fiorin, M. L. Klein, J. Hénin, *Mol. Phys.* **2013**, *111*, 3345.
[52] M. Sprik, *Chem. Phys.* **2000**, *258*, 139.
[53] M. Kılıç, B. Ensing, *Phys. Chem. Chem. Phys.* **2014**, *16*, 18993.
[54] M. Sulpizi, M. Sprik, *Phys. Chem. Chem. Phys.* **2008**, *10*, 5238.
[55] J. Cheng, M. Sulpizi, M. Sprik, *J. Chem. Phys.* **2009**, *131*, 154504.
[56] F. Costanzo, M. Sulpizi, R. G. Della Valle, M. Sprik, *J. Chem. Phys.* **2011**, *134*, 244508.
[57] J. Cheng, C. Liu, J. VandeVondele, M. Sulpizi, M. Sprik, *Acc. Chem. Res.* **2014**, *47*, 3522.
[58] R. A. Marcus, N. Sutin, *Biochim. Biophys. Acta Rev. Bioenerg.* **1985**, *811*, 265.
[59] G. King, A. Warshel, *J. Chem. Phys.* **1990**, *93*, 8682.
[60] M. Sulpizi, M. Sprik, *J. Phys. Condens. Matter* **2010**, *22*, 284116.
[61] X. Liu, X. Lu, M. Sprik, J. Cheng, E. J. Meijer, R. Wang, *Geochim. Cosmochim. Acta* **2013**, *117*, 180.
[62] T. De Meyer, K. Hemelsoet, L. Van Der Schueren, E. Pauwels, K. De Clerck, V. Van Speybroeck, *Chem. Eur. J.* **2012**, *18*, 8120.
[63] E. A. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chem. Phys. Lett.* **1989**, *156*, 472.
[64] W. K. Den Otter, W. J. Briels, *J. Chem. Phys.* **1998**, *109*, 4139.
[65] M. Sprik, *Faraday Discuss.* **1998**, *110*, 437.
[66] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello, *Comput. Phys. Commun.* **2009**, *180*, 1961.
[67] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604.

[68] P. L. Geissler, C. Dellago, D. Chandler, *J. Phys. Chem. B* **1999**, *103*, 3706.

[69] G. A. Tribello, M. Ceriotti, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5196.

[70] O. Valsson, P. Tiwary, M. Parrinello, *Annu. Rev. Phys. Chem.* **2016**, *67*, 159.

[71] Y. Matsunaga, Y. Komuro, C. Kobayashi, J. Jung, T. Mori, Y. Sugita, *J. Phys. Chem. Lett.* **2016**, *7*, 1446.

[72] *A Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Eds.: M. Abramowitz, I. Setegun), Dover Publications, New, York, **1965**.

[73] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, Cambridge, **2007**.

[74] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Comput. Phys. Commun.* **2005**, *167*, 103.

[75] G. Lippert, J. Hutter, M. Parrinello, *Mol. Phys.* **1997**, *92*, 477.

[76] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.

[77] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.

[78] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[79] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, *54*, 1703.

[80] J. VandeVondele, J. Hutter, *J. Chem. Phys.* **2007**, *127*, 114105.

[81] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.

[82] G. J. Martyna, M. L. Klein, M. Tuckerman, *J. Chem. Phys.* **1992**, *97*, 2635.

[83] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, M. Parrinello, *J. Chem. Phys.* **2005**, *122*, 014515.

[84] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

[85] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247.

[86] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, *J. Comput. Chem.* **2009**, *30(13)*, 2157.

[87] T. Verstraelen, L. Vanduyfhuys, S. Vandenbrande, S. M. J. Rogge, Yaff, *Yet Another Force Field*, **2016**; http://molmod.ugent.be/software/yaff.

[88] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.

[89] H. Flyvbjerg, H. G. Petersen, *J. Chem. Phys.* **1989**, *91*, 461.

[90] C. Y. Legault, *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, D. J. Fox, *Gaussian 09*, *Revision D*, Gaussian, Inc., Wallingford, CT, **2009**.

[92] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[93] E. Serjeant, B. Dempsey, *Ionisation Constants of Organic Acids in Aqueous Solution*, Pergamon, Oxford, **1979**.

[94] J. Drahoňovský, Z. Vacek, *Collect. Czech. Chem. Commun.* **1971**, *36*, 3431.

[95] G. Kortüm, W. Vogel, K. Andrussow, *Dissociation Constants of Organic Acids in Aqueous Solution*, Butterworth, London, **1961**.

[96] P. J. Pearce, R. J. J. Simkins, *Can. J. Chem.* **1968**, *46*, 241.