Absolute pK_a Values and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation

Martina Mangold,[†] Leslie Rolland,[‡] Francesca Costanzo,[§] Michiel Sprik,[†] Marialore Sulpizi,[†] and Jochen Blumberger*,[∥]

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[‡]Departement de Chimie, Ecole Normale Superieure, 24 rue Lhomond, 75231 Paris Cedex 05, France

[§]Dipartimento di Chimica Fisica e Inorganica, Universita di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy

[∥]Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom

Ⓢ *Supporting Information*

ABSTRACT: Absolute pK_a values of the amino acid side chains of arginine, aspartate, cysteine, histidine, and tyrosine; the C- and N-terminal group of tyrosine; and the tryptophan radical cation are calculated using a revised density functional based molecular dynamics simulation technique introduced previously [Cheng, J.; Sulpizi, M.; Sprik, M. *J. Chem. Phys.* **2009**, *131*, 154504]. In the revised scheme, acid deprotonation is considered as a dissociation rather than a proton transfer reaction, and a correction term for treating the proton as a hydronium ion is suggested. The acidity constants of the amino acids are obtained from the vertical energy gaps for removal or insertion of the acidic proton and the computed solvation free energy of the proton. The unsigned mean error relative to experimental results is 2.1 pK_a units with a maximum error of 4.0 pK_a units. The estimated mean statistical uncertainty due to the finite length of the trajectories is ±1.1 pK_a units. The solvation structures of the protonated and deprotonated amino acids are analyzed in terms of radial distribution functions, which can serve as reference data for future force field developments.

1. INTRODUCTION

The prediction of pK_a values of solvated molecules has attracted much attention in the computational chemistry community over many years.[1−32] This is not very surprising if one takes into consideration that proton transfer is the most frequently occurring reaction in nature, and an often encountered reaction in technological processes. Protons play an important role in energy conversion in living cells and fuel cells, facilitate ion exchange in biological and synthetic membranes, and catalyze chemical reactions at the active site of proteins and in synthetic reactions. Central to a quantitative characterization of such processes is the knowledge of the pK_a values of the molecules involved, as this quantity determines the protonation state of the system at a given pH as well as the energetics for intra- or intermolecular proton transfer. The ability to predict pK_a values from computation is very important, in particular when the system under consideration is not amenable to experimental measurements.

The majority of pK_a calculations of solvated molecules have been carried out by treating the solute at the QM level, while assuming that the interactions with the environment can be modeled by an electric continuum; see refs 1−18 for a selection of papers that have appeared in the past 10 years. Using thermodynamic cycles, the deprotonation free energy is expressed as the sum of the deprotonation free energy of the molecule in the gas phase and the solvation free energy difference of products (deprotonated acid plus proton) and the reactant (protonated acid). The advantage of such a scheme is that high-level *ab initio* methods can be used to describe the chemical deprotonation step. The disadvantage is that short-range

intermolecular interactions with solvent molecules, such as hydrogen bonds and ion—dipole interactions, are not explicitly accounted for, although they are considered to be particularly important to describe solvation free energies.

In a recent study, a number of QM-continuum protocols for calculation of pK_a values of biologically important carbonic acids have been compared.[16] While none of the standard continuum calculations were reported to give a satisfactory overall performance for neutral and charged acids, it was possible to obtain in some cases good agreement with experimental results by either comparing relative pK_a values of iso-Coulombic deprotonation reactions or by adding explicit solvent molecules in the QM calculation. It was concluded that the consideration of the "real" character of the solvent is of major importance in the future development of solvation models beyond the continuum approximation.

An alternative to QM-continuum computations is the density functional based molecular dynamics (DFMD) calculation of pK_a values.[25−32] In contrast to QM-continuum methods, in the DFMD approach, both the solute and the solvent are treated at the DFT level of theory, and the deprotonation free energy can be obtained from the statistical mechanics formalism of condensed phase simulation. The advantage of DFMD is that solute—solvent interactions are accounted for at the same QM level of theory, albeit at a higher computational cost and a more restricted choice of the electronic structure method. In early applications, proton transfer free energies were obtained from

the potential of mean force for transfer of a proton from the solvated acid to a neighboring water molecule.[25,26] Using this method, the pK_a value of an inorganic acid[25] could be calculated to a good degree of accuracy as well as the proton transfer free energy in an enzyme active site.[26]

More recently, some of us have developed an alternative simulation protocol, by combining DFMD with the thermodynamic integration method (DFMD-TI).[29−32] Here, the pK_a value is obtained in a two-step process. In the first step, the free energy is calculated for the alchemical transformation of the acidic proton into a dummy atom in the condensed phase by sampling the vertical DFT energy gap for this transformation along DFMD trajectories. In the second reaction step, the free energy for the transfer of the dummy atom into the gas phase is calculated. The absolute deprotonation free energy is the sum of the two contributions, and the pK_a value is the difference in the absolute deprotonation free energy for the acid under consideration and the solvated proton (see section 2 for details). Similar simulation approaches have been carried out before with classical[19−21] or QM/MM[22−24] potentials. However, application in the framework of all-QM DFMD remains scarce.

The aim of our present study is three-fold. First, we present a scheme for the DFMD-TI computation of pK_a values that is somewhat different from our previous formulation. Here, we consider deprotonation as a dissociation reaction rather than a proton transfer reaction, and we suggest a correction term for treating the solvated proton as a hydronium ion. Second, we investigate the performance of this method for the prediction of absolute pK_a values of naturally occurring amino acids. In particular, we investigate the relative error of this method with respect to experimental results, and how the achieved accuracy compares with that of typical continuum calculations. Third, as the pK_a calculation requires us to run DFMD of the protonated and deprotonated solute, we provide reference data for the solvation structure of the amino acids in both protonation states. Such structural information is important for the future development of QM/MM models for protein pK_a predictions, where aqueous amino acid solutions serve as reference states. The pK_a calculations carried out for tyrosine and the tryptophan cation radical are similar to the ones of ref 33, where proton transfer coupled reduction potentials of these amino acids are reported.

We find that DFMD-TI in the present form can predict absolute acidity constants of amino acids to an accuracy of ±2.1 pK_a units, which is comparable to the performance of some of the best QM-continuum calculations carried out recently for a set of similar acids.[17]

2. THEORY

2.1. Definition of pK_a Value.
We consider the dissociation of an acidic molecule AH in aqueous (aq) solution to the solvated proton $H^+(aq)$, and the conjugate base $A^-(aq)$,

$$AH(aq) \rightarrow A^-(aq) + H^+(aq) \qquad (1)$$

The pK_a value of AH(aq) is defined as the negative logarithm of the equilibrium constant K_a of reaction eq 1:

$$pK_a = -\log K_a \qquad (2)$$

$$K_a = \left[\frac{a_{A^-} a_{H^+}}{c^\circ a_{AH}} \right] \qquad (3)$$

where a_X is the equilibrium activity of species X, which is equal to the equilibrium concentration for dilute solutions, and $c^\circ = 1\ mol\ dm^{-3}$ is the standard concentration.

The aim is the computation of the standard reaction free enthalpy of reaction eq 1, $\Delta_{dp}G^\circ$, related to the equilibrium constant by

$$\Delta_a G^\circ = -k_B T \ln K_a \qquad (4)$$

where the subscript "dp" refers to deprotonation. In practice, we neglect the (small) difference between free enthalpies and free energies in condensed phase reactions and formulate a computational scheme for the reaction free energy for deprotonation, $\Delta_a A^\circ$. For this purpose, we split the full reaction eq 1 into two steps. In the first step, the proton is transferred into the gas phase while the conjugate base remains in aqueous solution:

$$AH(aq) \rightarrow A^-(aq) + H^+(g) \qquad (5)$$

and in the second step the proton is transferred from the gas to the solution phase, the reverse process of the reaction
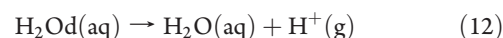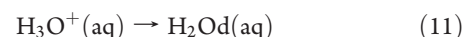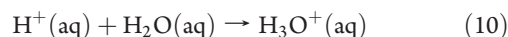
$$H^+(aq) \rightarrow H^+(g) \qquad (6)$$

In order to make the reaction eq 5 amenable to DFMD computation, we describe it by a series of two alchemical transformations:

$$AH(aq) \rightarrow Ad^-(aq) \qquad (7)$$

$$Ad^-(aq) \rightarrow A^-(aq) + H^+(g) \qquad (8)$$
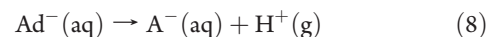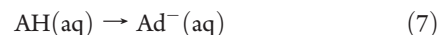
In the reactant state of reaction eq 7, the acidic proton is treated like the rest of the system, at the DFT level of theory. The proton is transformed into a dummy atom in the product state as indicated by the symbol "d". The dummy atom is connected to A by harmonic spring potentials (see eq 18), but it does not interact with the system otherwise. In reaction eq 8, the artificial spring is removed, and the dummy atom is transferred from the aqueous to the gas phase and transformed back into a proton. The free energy changes of reaction eqs 7 and 8 are denoted as $\Delta_{dp}A_{AH}$ and $-\Delta A_{Ad}$, respectively. When adding a quantum correction for nuclear motion, ΔA_{qc}, one obtains the reaction free energy of reaction eq 5:

$$\Delta_{dp}A_{AH}^\circ = \Delta_{dp}A_{AH} - \Delta A_{Ad} - \Delta A_{qc} \qquad (9)$$

The computation of the desolvation free energy of the aqueous proton, reaction eq 6, is complicated by the transient nature of this species. The solvated proton is often modeled as a hydronium ion ($H_3O^+(aq)$), but one should bear in mind that this is only an idealized structure adopted by the proton during its structural diffusion in aqueous solution.[34] Thus, we make a clear distinction between the "real" solvated proton $H^+(aq)$ and the idealized hydronium ion. In order to compute the desolvation free energy, we consider the following reaction steps:

$$H^+(aq) + H_2O(aq) \rightarrow H_3O^+(aq) \qquad (10)$$

$$H_3O^+(aq) \rightarrow H_2Od(aq) \qquad (11)$$

$$H_2Od(aq) \rightarrow H_2O(aq) + H^+(g) \qquad (12)$$

In the first step, reaction eq 10, the idealized hydronium ion is "assembled" from the hydrated proton and a water molecule with a reaction free energy of $-\Delta A_{H_3O^+}$. The second and the third

reaction steps, eqs 11 and 12 with reaction free energies $\Delta_{dp}A_{AH}$ and $-\Delta A_{H_2Od}$, respectively, are similar to the two reaction steps eqs 7 and 8 with AH replaced by the hydronium ion. The sum of all free energy contributions gives the free energy for desolvation of the aqueous proton:

$$\Delta_{ds}A_{H^+}^{\circ} = \Delta_{dp}A_{H_3O^+} - \Delta A_{H_2Od} - \Delta A_{qc} - \Delta A_{H_3O^+} \quad (13)$$

where we have also added a quantum correction term.

Combining eqs 2, 4, 9, and 13, the pK_a value of AH is given by the difference in the deprotonation free energy of AH and the desolvation free energy of the solvated proton, reaction eqs 5 and 6:

$$pK_a = \frac{1}{\ln 10 k_B T}(\Delta_{dp}A_{AH}^{\circ} - \Delta_{ds}A_{H^+}^{\circ}) \quad (14)$$

with $\Delta_{dp}A_{AH}^{\circ}$ and $\Delta_{ds}A_{H^+}^{\circ}$ computed according to eqs 9 and 13, respectively.

The present definition of pK_a is somewhat different from the one used previously.[29−31] Here, we consider acid deprotonation according to Arrhenius as a *dissociation* reaction (eq 1), whereas in previous work we considered it according to Brønsted, as a *proton transfer* reaction:

$$AH(aq) + H_2O(l) \rightarrow A^-(aq) + H_3O^+(aq) \quad (15)$$
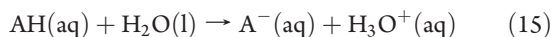
with equilibrium constant

$$K_a' = \left[\frac{a_{A^-} a_{H_3O^+}}{c^{\circ} a_{AH}}\right] \quad (16)$$

The subtle difference between the two definitions is that in the Brønsted picture one assumes that the acidic proton is chemically bonded to a solvent molecule forming a solvated hydronium ion, whereas in the Arrhenius picture this assumption is not made. Indeed, the equilibrium constants eqs 3 and 16 differ only by the notation for the proton, a_{H^+} and $a_{H_3O^+}$, respectively. While this distinction is of course irrelevant in experiments, it affects the results of computations, where a model for the proton in solution must be chosen. The difference between the two pK_a definitions is the free energy for assembling the hydronium ion from the solvated proton and a water molecule, $-\Delta A_{H_3O^+}$ of reaction eq 10. We will estimate $\Delta A_{H_3O^+}$ in section 2.5, after we have described the computation of the other free energy terms that appear in the expressions for $\Delta_{dp}A_{AH}^{\circ}$ and $\Delta_{ds}A_{H^+}^{\circ}$.

2.2. Free Energy for Alchemical Transformation, $\Delta_{dp}A_{AH}$. The free energy for transformation of the proton into a dummy atom, $\Delta_{dp}A_{AH}$ of reaction eqs 7 and 11, is computed using Kirkwood's coupling parameter method.[35] A mapping potential E_η is defined that couples the potential energy surfaces of the reactant and product states.

$$E_\eta = (1 - \eta)E_0 + \eta E_1 \quad (17)$$

In eq 17, η is a coupling parameter that takes values from 0 to 1. E_0 is the DFT potential energy surface (PES) of AH(aq) (or H_3O^+(aq)), and E_1 is the DFT PES of Ad$^-$(aq) (or H_2Od(aq)) plus the restraining potential V_r that keeps the dummy atom attached to A (H_2O).

$$V_r = \sum_{bonds} \frac{k_r}{2}(r - r_{eq})^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_{eq})^2$$
$$+ \sum_{dihedrals} \frac{k_\varphi}{2}(\varphi - \varphi_{eq})^2 \quad (18)$$

The presence of the restraining potential in the product state E_1 allows one to generate positions for insertion of the proton as required for the energy gap calculation according to eq 20. Note that E_0 of the weakly acidic amino acids does not contain the restraining potential V_r, whereas E_0 of the hydronium ion does contain V_r in order to suppress spontaneous dissociation and proton hopping that would otherwise occur during the DFMD simulation.

The free energy difference between reactant and product described by the PESs E_0 and E_1, respectively, is the thermo-dynamic integral:

$$\Delta_{dp}A_{AH} = A(\eta = 1) - A(\eta = 0)$$
$$= \int_0^1 d\eta \langle\Delta_{dp}E_{AH}\rangle_\eta \quad (19)$$

where ΔE is the vertical energy gap

$$\Delta_{dp}E_{AH} = E_1 - E_0 \quad (20)$$

and the brackets $\langle\cdots\rangle_\eta$ denote the usual thermal average on the PES E_η. The vertical energy gap eq 20 is sampled along DFMD trajectories on the PES E_η. In present calculations, the full thermodynamic integral eq 19 is approximated using Simpson's rule:

$$\Delta_{dp}A_{AH} \approx \Delta A_{TP}$$
$$= \frac{1}{6}(\langle\Delta_{dp}E_{AH}\rangle_0 + \langle\Delta_{dp}E_{AH}\rangle_1) + \frac{2}{3}\langle\Delta_{dp}E_{AH}\rangle_{0.5} \quad (21)$$

and compared to the standard linear response (LR) expression:

$$\Delta_{dp}A_{AH} \approx \Delta A_{LR} = \frac{1}{2}(\langle\Delta_{dp}E_{AH}\rangle_0 + \langle\Delta_{dp}E_{AH}\rangle_1) \quad (22)$$

The LR approximation is exact if the thermal probability distribution of ΔE is Gaussian, or equivalently, if the root-mean-square fluctuations (rmsf) of the energy gap

$$\sigma_\eta = (\langle(\Delta_{dp}E_{AH} - \langle\Delta_{dp}E_{AH}\rangle_\eta)^2\rangle_\eta)^{1/2} \quad (23)$$

are the same for any value of η.

2.3. Free Energy for Dummy Atom Insertion, ΔA_{Ad}. The reaction free energy for the reverse of reactions eq 8 and 12, ΔA_{Ad}, is related to the restraining potential V_r by

$$\Delta A_{Ad} = k_B T \ln\langle\exp(\beta V_r)\rangle_1 \quad (24)$$

Equation 24 is rigorous and can be readily derived from statistical mechanical principles (see the Supporting Information (SI) of this article for a proof). It is the free energy of transforming a free dummy particle confined to the volume $V^\circ = 1$ mol/c° plus a solvated species A$^-$ at concentration c° and described by the PES $E_1 - V_r$, to the solvated species Ad$^-$, described by the PES E_1 (that is, with the dummy atom attached to A$^-$, see eq 12 in the SI). In the derivation of eq 24, the standard concentrations of the aqueous phase and gas phase are assumed to be the same, $c^\circ = 1$ mol dm^{-3}. A term $k_B T \ln(c^\circ RT/p^\circ)$ has to be added to eq 24 if the conventional standard state of the gas phase is used. The free energy ΔA_{Ad} thus depends on the strength of the dummy potential, and also on the definition of the standard volume V°.

Our previously derived expression for ΔA_{Ad} reads[30]

$$\Delta A_{Ad} = -k_B T \ln(c^\circ \Lambda_{H^+}^3 f_{Ad}) \quad (25)$$

where f_{Ad} is the ratio of the internal gas-phase partition functions of Ad^- and A^-; $f_{Ad} = q_{Ad^-}^{int}/q_{A^-}^{int}$ with q^{int} the product of electronic, vibrational, and rotational partition functions; and Λ_{H^+} is the thermal wavelength of the proton (1.01 Å at 298 K). In the SI, we show that eq 25 follows from eq 24, if one assumes that the excess chemical potentials of A^- and Ad^- as well as their thermal wavelengths are the same. One can expect that the first approximation is a very good one as long as the presence of the dummy atom does not induce any major changes in the equilibrium structure of the solute. The second approximation is valid if the mass of A is much larger than that of the dummy atom, which is usually the case. As discussed in the SI, we prefer to calculate ΔA_{Ad} according to eq 25, since the MD estimate of the expectation value on the rhs of eq 24 gives very poor results. The ratio of gas-phase partition functions, f_{Ad}, is approximated by the three mode vibrational partition function:[30]

$$f_{Ad} = \frac{T^3}{\theta_1^{vib}\theta_2^{vib}\theta_3^{vib}} \qquad (26)$$

where $\theta_i^{vib} = h\nu_{Ad,i}/k_B$ is the vibrational temperature and $\nu_{Ad,i}$ is the frequency of mode i. The approximation eq 26 is justified when the mass of A is sufficiently large compared to the mass of the dummy atom, so that the vibrations of the latter can be treated as local modes decoupled from the molecular modes of A. The effect of the dummy atom on the inertial tensor used to calculate the rotational partition function is ignored. We have previously shown that under these approximations ΔA_{Ad} is the free energy for confining the free dummy particle from volume V° to the volume $(2\pi)^{3/2}\Delta_1\Delta_2\Delta_3$, where $\Delta_i = (k_BT/k_i)^{1/2}$ is the mean square displacement of the local normal mode i of the dummy particle with force constant k_i.[30] The same physically appealing picture can be derived from eq 24 (see SI for details). For f_{H_2Od}, eq 26 is not a good approximation. Here, the full classical partition functions of H_2Od and H_2O have to be calculated. We refer to ref 30 for an explicit expression.

2.4. Quantum Corrections, ΔA_{qc}. The first two terms on the rhs of eqs 9 and 13, $\Delta_{dp}A_{AH} - \Delta A_{Ad}$, are the reaction free energy for the deprotonation reactions eqs 7 and 8 and eqs 11 and 12, if one assumes classical ionic motion. Nuclear quantum effects are expected to be significant, however, because in these reactions three modes of rather high frequency are lost when lifting the proton into the gas phase. We correct for this effect by subtracting the classical vibrational free energy of the acid in the gas phase using the three mode model (ΔA_{cl}^{vib}) and adding the corresponding quantized vibrational free energy (ΔA_q^{vib}):

$$\Delta A_{qc}(X) = \Delta A_q^{vib}(X) - \Delta A_{cl}^{vib}(X) \qquad (27)$$

For $X = AH$,

$$\Delta A_q^{vib}(AH) = -k_BT \ln \prod_{i=1}^{3} \frac{\exp(-h\nu_{AH,i}/(2k_BT))}{1 - \exp(-h\nu_{AH,i}/(k_BT))}$$

$$\approx \sum_{i=1}^{3} \frac{1}{2}h\nu_{AH,i} \qquad (28)$$

$$\Delta A_{cl}^{vib}(AH) = -k_BT \ln \prod_{i=1}^{3} \frac{k_BT}{h\nu_{AH,i}} \qquad (29)$$

where $\nu_{AH,i}$ denotes the three modes of the proton. For the hydronium ion, $X = H_3O^+$, all modes of the solute are taken into account:

$$\Delta A_q^{vib}(H_3O^+) = \sum_{i=1}^{6} \frac{1}{2}h\nu_{H_3O^+,i} - \sum_{i=1}^{3} \frac{1}{2}h\nu_{H_2O,i} \qquad (30)$$

$$\Delta A_{cl}^{vib}(H_3O^+) = -k_BT\left(\ln \prod_{i=1}^{6} \frac{k_BT}{h\nu_{H_3O^+,i}} - \ln \prod_{i=1}^{3} \frac{k_BT}{h\nu_{H_2O,i}}\right) \qquad (31)$$

2.5. Dissociation Free Energy of H_3O^+, $\Delta A_{H_3O^+}$. The free energy for dissociation of the hydronium ion into the solvated proton and a water molecule, the reverse process of reaction eq 10, can be written as follows:

$$\Delta A_{H_3O^+} = k_BT \ln \frac{c^\circ \Lambda_{H^+}^3 \Lambda_{H_2O}^3}{\Lambda_{H_3O^+}^3} \frac{q_{H_3O^+}^q}{q_{H^+}^q q_{H_2O}^q} \qquad (32)$$

where q_X^q is the quantum partition function of the solvated species X after separation of the thermal wavelength Λ_X. The ratio of partition functions on the rhs of eq 32 is difficult to calculate in practice due to the delocalized nature of the solvated proton. To estimate this term, we adopt a mixed quantum-classical scheme, where the hydronium ion is modeled with one quantum proton and two classical protons, $H_2OH_q^+$, interacting with a classical environment. The ratio of partition functions $q_{H_3O^+}^q/q_{H_2O}^q$ is then approximated by an effective quantum vibrational partition function of the single quantum proton, $q_{H^+,eff}^q = q_{H_3O^+}^q/q_{H_2O}^q$:

$$\Delta A_{H_3O^+} = k_BT \ln c^\circ \Lambda_{H^+}^3 \frac{q_{H^+,eff}^q}{q_{H^+}^q}$$

$$= k_BT \ln c^\circ \Lambda_{H^+}^3 \frac{q_{H^+,eff}^{q'}}{q_{H^+}^q} - \Delta E_{H^+,eff}^{zp} \qquad (33)$$

where we have neglected the (small) difference in the thermal wavelengths of water and the hydronium ion, similarly as in eq 25. In the last equation, we have separated $q_{H^+,eff}^q$ into the zero-point energy term $\Delta E_{H^+,eff}^{zp}$ and the partition function $q_{H^+,eff}^{q'}$ in which vibrational energies are referred to the vibrational ground state, so that $q_{H^+,eff}^{q'} = 1$ for $T = 0$. Assuming that the zero-point motion for the proton in the hydronium ion and the solvated proton are similar, we obtain

$$\Delta A_{H_3O^+} \approx k_BT \ln c^\circ \Lambda_{H^+}^3 q_{H^+,eff}^{q'} \qquad (34)$$

Equation 34 is of course an approximation that could in principle be validated with path-integral DFMD calculations. This equation can be further simplified by noting that the vibrational temperature constant of a proton is significantly higher than room temperature, i.e., $q_{H^+,eff}^{q'} \approx 1$. Thus, under the above assumptions, the dissociation free energy is equal to the translational free energy of the proton

$$\Delta A_{H_3O^+} \approx k_BT \ln c^\circ \Lambda_{H^+}^3 \qquad (35)$$

We find that eq 35 compensates for the loss in the translational free energy for dummy atom insertion (eq 25), thereby leaving the

1954

dx.doi.org/10.1021/ct100715x |*J. Chem. Theory Comput.* 2011, 7, 1951–1961

Table 1. Force Constants and Equilibrium Values for Bonds (k_r, r_{eq}), Angles (k_θ, θ_{eq}), and Dihedrals (k_φ, φ_{eq}) of the Restraining Potential V_r for the Dummy Particle (eq 18)

	k_r (au)	r_{eq} (Å)	k_θ (au)	θ_{eq} (deg)	k_φ (au)	φ_{eq} (deg)	n (H_2O)[a]
Arg_{mod}	0.2	1.03	0.2	122	0.02	13	54
Asp	0.2	1.02	0.2	111	0.20	9	53
Cys	0.2	1.35	0.2	97	0.02	92	55
His	0.2	1.05	0.2	123	0.02	173	54
Trp*	0.2	1.03	0.2	125	0.20	178	48
Tyr	0.2	1.00	0.2	110	0.02	20	53
Tyr_C	0.2	1.03	0.2	112	0.02	166	53
Tyr_N	0.2	1.05	0.2	112	0.02	47	53
H_3O^+	0.2	1.00	0.2	111			63

[a] Number of solvent water molecules per unit cell.

number of degrees of freedom unchanged. Indeed, this should be the case for the transfer of a proton from the gas to the aqueous phase.

2.6. Final Expression for pK_a. We can now write the pK_a explicitly in terms of the deprotonation integrals and the dummy, quantum, and dissociation corrections. Insertion of eqs 9 and 13 in eq 14 gives

$$pK_a = \frac{1}{\ln 10 k_B T}(\Delta_{dp}A_{AH} - \Delta_{dp}A_{H_3O^+} - \Delta A_{Ad} + \Delta A_{H_2Od}$$
$$- \Delta A_{qc}(AH) + \Delta A_{qc}(H_3O^+) + \Delta A_{H_3O^+}) \quad (36)$$

and insertion of eqs 19, 25, and 35 gives

$$pK_a = \frac{1}{\ln 10 k_B T}\left(\int_0^1 d\eta \langle \Delta_{dp}E_{AH}\rangle_\eta - \int_0^1 d\eta \langle \Delta_{dp}E_{H_3O^+}\rangle_\eta \right.$$
$$+ k_B T \ln \frac{q_{Ad^-}^{int} q_{H_2O}^{int}}{q_{A^-}^{int} q_{H_2Od}^{int}} - \Delta A_{qc}(AH) + \Delta A_{qc}(H_3O^+)$$
$$\left. + k_B T \ln c^\circ \Lambda_{H^+}^3 \right) \quad (37)$$

We will see that the vibrational frequencies of the dummy atom in the acid (Ad$^-$) and in the hydronium (H$_2$Od) are rather similar. Consequently, the third term on the rhs of eq 37 will be small. The same is true for the proton frequencies in most acids, leading to partial cancellation of the fourth and fifth terms on the rhs of eq 37. Thus, the pK_a of an acid is approximately equal to the difference of the deprotonation integral of the acid and the hydronium ion corrected by the standard translational free energy of the proton.

3. SIMULATION DETAILS

Aqueous solutions of amino acids were simulated under periodic boundary conditions. Each unit cell comprised of one amino acid molecule (modified arginine (Arg$_{mod}$), aspartic acid (Asp), cysteine (Cys), histidine (His), tryptophan radical cation (Trp*), and tyrosine (Tyr)), and a variable number of water molecules (48−55, see Table 1). A model aqueous solution for the simulation of Tyr is shown in Figure 1. The long side chain of arginine was shortened by two CH$_2$ groups in order to fit the molecule into the simulation box specified below. We refer to this amino acid as Arg$_{mod}$. The pK_a calculations are for the side chain ionizable proton, if not specified otherwise, and the N- and C-termini of the amino acids are treated in the ionized form. For pK_a calculation of the C-terminus of tyrosine (Tyr$_C$), the
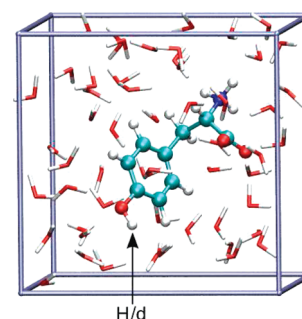
Figure 1. Aqueous model solution used for calculation of the pK_a value of tyrosine (Tyr). The latter is depicted in ball and stick representation, and the water molecules are depicted in stick representation. Color code: C, cyan; O, red; N, blue; H, white. The arrow indicates the hydrogen atom that is treated quantum mechanically in the protonated state (H) and as a dummy atom in the deprotonated state (d).

Table 2. Local Normal Mode Frequencies of the Dummy Atom d ($\nu_{Ad,i}$) and of the Quantum Mechanically Treated Proton H ($\nu_{AH,i}$) for the Gas Phase Amino Acids Ad and AH, Respectively, and the Vibrational Temperature θ_i^{vib} Corresponding to $\nu_{Ad,i}$

	$\nu_{Ad,i}$ (cm^{-1})	θ_i^{vib} (K)	$\nu_{AH,i}$ (cm^{-1})
Arg_{mod}	440, 1182, 2290	633, 1701, 3295	263, 1204, 3340
Asp	1186, 1797, 2290	1706, 2585, 3295	577, 1141, 3446
Cys	284, 893, 2290	409, 1285, 3295	240, 877, 2542
His	433, 1154, 2290	623, 1660, 3295	575, 1165, 3417
Trp*	1182, 1448, 2290	1701, 2083, 3295	503, 1203, 3392
Tyr	1212, 1296, 2290	1744, 1865, 3295	382, 1208, 3566
Tyr_C	396, 1171, 2290	570, 1685, 3295	667, 1131, 3420
Tyr_N	384, 1152, 2290	552, 1657, 3295	1183, 1227, 3173
H_3O^+ [a]	1450, 1810, 1920	2090, 2600, 2760	1032, 1668, 1692
	3340, 3660, 3760	4810, 5260, 5400	3449, 3523, 3540

[a] $\nu_{Ad,i}$ and θ_i^{vib} from ref 30.

N-terminus and the side chain were protonated, and for pK_a calculation of the N-terminus (Tyr$_N$), the C-terminus was deprotonated and the side chain was protonated. The net charge of the solutions is compensated by the neutralizing background charge included in the Ewald summation of the electrostatic energy. The number of water molecules per unit cell was obtained from classical MD simulation runs of each amino acid, by varying the number of solvent molecules until a box length close to 12.426 Å was obtained. This value was taken as the dimension of the cubic unit cell in all DFMD simulations described below.

A snapshot from an equilibrated classical MD trajectory of the protonated amino acid ($\eta = 0$) was taken as an initial configuration for DFMD. The solution was equilibrated for 5 ps at 330 K, and the following 10−15 ps of dynamics was taken for calculation of thermal averages. A higher temperature than the standard temperature of 298 K is chosen to compensate for the underestimation of dynamical properties due to the approximate nature of the exchange correlation functional used (see below). The MD time step was 0.5 fs. Trajectories for the deprotonated state ($\eta = 1$) and for the midpoint ($\eta = 0.5$) were generated similarly. The thermal averages in eqs 21 and 22 were obtained by calculating the energy gap, eq 20, for about 150 equidistantly spaced snapshots of these trajectories. The calculations for

1955

dx.doi.org/10.1021/ct100715x |*J. Chem. Theory Comput.* 2011, 7, 1951–1961

Table 3. Summary of Vertical Energy Gaps and Root-Mean-Square Fluctuations for Removal and Insertion of an Excess Proton out of or into Aqueous Solutions of Amino Acids, As Obtained from Density Functional Based Molecular Dynamics Simulation[a]

AH	$\langle\Delta_{dp}E_{AH}\rangle_0$[b]	σ_0[c]	$\langle\Delta_{dp}E_{AH}\rangle_{0.5}$[b]	$\sigma_{0.5}$[c]	$\langle\Delta_{dp}E_{AH}\rangle_1$[b]	σ_1[c]	ΔA_{LR}[d]	ΔA_{TP}[e]	ΔA_{Ad}[f]	ΔA_{qc}[g]
Arg_{mod}	17.75 ± 0.11	0.29	16.48 ± 0.03	0.26	14.08 ± 0.13	0.84	15.92	16.29	0.32	0.17
Asp	17.43 ± 0.02	0.30	15.55 ± 0.10	0.28	13.41 ± 0.10	0.65	15.42	15.50	0.35	0.18
Cys	17.31 ± 0.08	0.26	15.97 ± 0.16	0.27	14.20 ± 0.06	0.49	15.79	15.93	0.33	0.12
His	17.33 ± 0.02	0.20	16.17 ± 0.02	0.21	13.14 ± 0.11	1.15	15.23	15.85	0.31	0.18
Trp*	16.97 ± 0.04	0.21	15.73 ± 0.03	0.19	13.60 ± 0.05	0.98	15.29	15.57	0.35	0.17
Tyr	17.79 ± 0.07	0.25	16.27 ± 0.13	0.33	12.67 ± 0.34	1.76	15.22	15.94	0.34	0.18
Tyr_C	17.40 ± 0.07	0.27	15.72 ± 0.08	0.30	13.61 ± 0.16	0.77	15.50	15.64	0.31	0.18
Tyr_N	17.49 ± 0.07	0.27	16.34 ± 0.10	0.26	13.06 ± 0.19	1.28	15.28	15.98	0.31	0.19
H_3O^+	17.40 ± 0.03	0.28	15.46 ± 0.00	0.31	11.98 ± 0.07	0.84	14.69	15.20	0.34	0.20

[a] Energies are for protonation/deprotonation of the side chain, except for Tyr_C and Tyr_N, where the C-terminus and N-terminus is protonated/deprotonated. All energies are in eV. [b] Thermal average of the energy gap eq 20 for $\eta = 0$, 0.5, and 1. [c] Equation 23 for $\eta = 0$, 0.5, and 1. [d] Equation 22. [e] Equation 21. [f] Equation 25 using the data given in Table 2. [g] Equation 27 using the data given in Table 2.

aqueous H_3O^+ were carried out in the Eigentype structure of the hydronium ion as described in ref 29. All three O—H bonds were restrained by a harmonic potential with a force constant of 1.0 au and an equilibrium bond length of 1.0 Å to avoid spontaneous proton transfer to a neighboring water molecule. The energy E_0 used to calculate the energy gap, eq 20, does not include these additional potential terms. The parameters defined in eq 18 for the dummy potential are summarized in Table 1. The frequencies used for the calculation of ΔA_{Ad} and ΔA_{qc} are summarized in Table 2.

The exchange-correlation energy was calculated according to Becke[36] and Lee, Yang, and Parr.[37] Core electrons and the nuclei were described using Goedecker−Teter−Hutter pseudopotentials, and the valence electrons were expanded in the TZV2P atomic basis set. The density cutoff for the auxiliary plane wave basis was 280 Ry. The tryptophan radical cation (Trp*) was treated at the local spin density functional level of theory, enforcing the lowest (doublet) spin state in both protonation states. DFMD simulations were carried out with the CP2k simulation package.[38] The initial equilibration of the structures was performed using the AMBER99 force field[39] and TIP3P water[40] as implemented in the NAMD simulation package.[41] Snapshots of the equilibrated systems can be made available on request.

4. RESULTS

4.1. Vertical Energy Gaps. The thermal average of the vertical energy gaps, $\langle\Delta_{dp}E_{AH}\rangle_\eta$, and the corresponding root-mean-square fluctuations (rmsf), σ_η, were obtained from DFMD simulation of the aqueous amino acids and are summarized in Table 3. The mean values are reasonably well converged, despite the rather short simulation time of 10−15 ps. The statistical uncertainty, estimated as half the difference of the mean value obtained for the first and the second half of the trajectory, is no more than about 0.1 eV for the initial ($\eta = 0$) and midpoint states ($\eta = 0.5$). The uncertainty for the final state ($\eta = 1$) is in general larger, 0.1−0.3 eV. This trend is paralleled by the rmsf of the energy gap. In the initial state, the rmsf is between 0.2 and 0.3 eV, which, interestingly, is in the same range as the rmsf reported for vertical ionization of aqueous solutes[42,43] and proteins.[44] The rmsf in the final state is significantly larger by a factor of 2 (Cys) to 7 (Tyr).

Table 4. Calculated Free Energies of Deprotonation and Comparison of Calculated and Experimental Absolute pK_a Values of Amino Acids

AH	$\Delta_{dp}A^\circ_{AH}$[a] (eV)	$\Delta_{dp}A^\circ_{AH}$[b] (eV)	pK_a (calc)[c]	pK_a (exp)[d]
Arg_{mod}	15.43 ± 0.08	15.80 ± 0.03	16.1 ± 0.6	12.10
Asp	14.89 ± 0.05	14.97 ± 0.07	2.0 ± 1.2	3.71
Cys	15.34 ± 0.05	15.48 ± 0.11	10.7 ± 1.8	8.14
His	14.74 ± 0.06	15.36 ± 0.02	8.6 ± 0.4	6.04
Trp*[e]	14.77 ± 0.03	15.05 ± 0.02	3.4 ± 0.4	4.30[f]
Tyr[e]	14.70 ± 0.17	15.42 ± 0.11	9.7 ± 1.8	10.10
Tyr_C	15.01 ± 0.09	15.15 ± 0.06	5.1 ± 1.0	2.24
Tyr_N	14.78 ± 0.10	15.48 ± 0.07	10.7 ± 1.2	9.04
H^+	14.34 ± 0.04	14.85 ± 0.01		

[a] Equation 9 for the amino acids, using the linear response approximation (LR) to $\Delta_{dp}A_{AH}$, ΔA_{LR} of eq 22, and eqs 25 and 27 for estimation of ΔA_{Ad} and ΔA_{qc}, respectively. For the solvated proton (H^+), the computed desolvation free energy is given, $\Delta_{ds}A^\circ_{H^+}$ of eq 13. ΔA_{H3O+} in eq 13 is approximated by eq 35. Data are taken from Table 3. [b] As in footnote a, but with $\Delta_{dp}A_{AH}$ estimated by the three-point approximation ΔA_{TP} of eq 21 (TP). [c] Equation 14 using the TP estimate for $\Delta_{dp}A_{AH}$. [d] Ref 48 if not specified otherwise. [e] The numerical values differ somewhat from the values given in ref 33 due to different system composition for Tyr and the neglect of quantum corrections for Tyr and Trp* in ref 33. [f] Ref 49.

The large fluctuations in the final state are a consequence of the very small distance between solvent molecules and the inserted proton. The water molecules are oriented so as to solvate the deprotonated acid Ad^- (note that the dummy atom does not interact with the solvent). Thus, two to three hydrogen atoms of the solvent are within 0.5−1 Å of the dummy atom. This unusually small H—H distance leads to relatively high energies when the dummy atom is transformed into the proton, as required for the calculation of the vertical energy gap, eq 20. As a consequence, the fluctuations of this energy and of the energy gap are large, too. Yet the distance is large enough for the insertion to be feasible; i.e., there is no hard core repulsion problem, which is usually the case in alchemical transformations. In fact, the proton affinities are still positive in the final state. By contrast, in the protonated state AH, the proton is typically solvated by one solvent molecule at a H—O distance of about 2 Å. Transformation of the proton into the dummy atom does not

1956

dx.doi.org/10.1021/ct100715x |J. Chem. Theory Comput. 2011, 7, 1951–1961

lead to high energy solvation structures. Hence, the fluctuations are smaller in the initial state.

According to our explanation, one would expect the difference in the rmsf of the gap in the initial and final states to increase with increasing strength of solvation of the deprotonated state (as this would lead to shorter H—H distances in the protonated state). This is indeed the case. For deprotonated Cys, the first peak of the radial distribution function between the S atom and the hydrogen atoms of the solvent (H_w) is at a rather large distance, 2.4 Å, giving the smallest difference in rmsf, whereas for deprotonated Tyr the first peak of the O—H_w radial distribution is at a very small distance, 1.6 Å, giving the largest difference in rmsf.

4.2. Free Energies and Absolute pK_a Values. The free energies $\Delta_{dp}A^\circ_{AH}$ of eq 9 and $\Delta_{ds}A^\circ_{H^+}$ of eq 13 are summarized in Table 4. They include the deprotonation free energy, the dummy and quantum correction, and, in case of $\Delta_{ds}A^\circ_{H^+}$, the thermochemical correction for the hydronium ion. The deprotonation free energy $\Delta_{dp}A_{AH}$ was calculated from the average energy gaps according to the three-point (TP) approximation of the thermodynamic integral, eq 21 (third column in Table 4), and according to the linear response formula (LR), eq 22 (second column in Table 4). It is already clear from the above observations that linear response is not a good approximation in the present calculations. For this approximation to hold, the rmsf's in the initial and final states have to be the same or should at least be similar. Indeed, the LR estimate differs by up to 0.7 eV (Tyr, Tyr_N) from the TP estimate, but also smaller deviations are observed (0.1 eV, Asp). Similar results were found in previous pK_a calculations of ammonia,[29] indicating that three points are the absolute minimum for evaluation of the thermodynamic integral, eq 19. An alternative cumulant expansion of the free energy $\Delta_{dp}A_{AH}$ up to second order (which requires the simulation of the two end states only, similarly to LR) does not give an improvement over the linear response estimate.

The deprotonation free energies $\Delta_{dp}A^\circ_{AH}$ obtained from the TP approximation of the thermodynamic integral correlate reasonably well with the experimental pK_a values, see Figure 2. The R^2 value for the linear fit is 0.93, and the computed slope of 64 meV is in good agreement with the theoretical slope of $k_BT \ln 10 = 59$ meV. The correlation with experimental results is poor for the linear response estimates, showing again that this approximation cannot be applied for the calculation of pK_a values. The absolute pK_a values are obtained according to eq 14 by subtracting the desolvation free energy of the solvated proton, $\Delta_{ds}A^\circ_{H^+}$, from $\Delta_{dp}A^\circ_{AH}$. The results are summarized in Table 4. We obtain an unsigned mean error (UME) of 2.1 pK_a units, a signed mean error (SME) of +1.3 pK_a units, and a maximum error of 4.0 pK_a units. The agreement is fairly good if one takes into account the average statistical uncertainty of 1.1 pK_a units, caused by the limited length of the DFMD trajectories.

4.3. Importance of $\Delta A_{H_3O^+}$, ΔA_{Ad}, and ΔA_{qc}. In view of possible simplifications of the computational procedure, it is of particular interest to analyze the importance of the terms $\Delta A_{H_3O^+}$, ΔA_{Ad}, and ΔA_{qc}, which relate the thermodynamic integral (eq 19) to the deprotonation free energies (eq 9) and absolute pK_a values (eq 2). Of the three corrections, the thermochemical correction for the hydronium ion, $\Delta A_{H_3O^+}$, is the most important one, amounting to −0.19 eV according to the approximation eq 35. The effect is a decrease in absolute pK_a by 3.2 units for all amino acids; relative pK_a values remain unchanged.
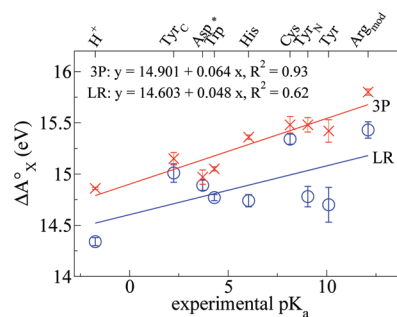

Figure 2. Computed free energies for deprotonation of amino acids in aqueous solution as obtained from density functional based molecular dynamics simulation, versus experimental pK_a values. The free energies $\Delta_{dp}A_{AH}$, defined in eq 9 for amino acids and in eq 13 for the solvated proton, as well as their error bars are taken from Table 4. Best linear fit functions for the free energies obtained from the three-point (×, TP, eq 21) and linear response (O, LR, eq 22) approximation to the thermodynamic integral, eq 19, are shown in red and blue, respectively.

Numerical values for the other terms ΔA_{Ad} and ΔA_{qc} are summarized in Table 3. We find that the dummy correction ΔA_{Ad} obtained from the gas phase partition function formula (eq 25) lies in the interval 0.31−0.35 eV for all amino acids investigated as well as for the hydronium ion. Thus, the effect of this correction is small, leading typically to an increase in absolute pK_a by less than 0.5 units and to a change in relative pK_a by less than 0.7 units. Indeed, we aimed at keeping the variation in ΔA_{Ad} small by choosing the equilibrium geometry of the dummy atom to be close to that of the proton. In this way, one also avoids introducing additional barriers for the transformation of the proton into the dummy atom.

The vibrational quantum correction ΔA_{qc} varies only marginally among amino acids, 0.17−0.19 eV, except for Cys, where a significantly smaller value of 0.12 eV is obtained. This trend is not surprising if one bears in mind that the quantum correction increases with increasing frequencies. While in all amino acids except Cys the proton is bound to a (light) second row atom (O or N), which gives rise to high frequencies in the range 3100−3600 cm^{-1}, in the case of Cys, the proton is bound to the heavier sulfur atom with a maximum frequency of about 2500 cm^{-1} (see Table 2). The hydronium ion has three modes that have a frequency larger than 3400 cm^{-1}. It thus exhibits the largest quantum correction. However, the difference in quantum correction between the amino acids and the hydronium ion (which enters the pK_a calculation) is rather small, 0.01−0.03 eV, but significant for Cys, 0.08 eV. This corresponds to an increase in the absolute pK_a value by 0.2−0.5 and 1.4 units, respectively. Thus, we conclude that the quantum correction for deprotonation from N and O atoms is relatively small, at least according to the simple correction model used, but can be significant if deprotonation from heavier heteroatoms is considered.

4.4. Solvation Structure. The solvation structure of the protonated and deprotonated amino acids is investigated to characterize the response of the solvent to the removal of the acidic proton. For this purpose, we have calculated the radial distribution function between the hydrogen atoms of solvent water molecules (H_W) and the heteroatom of the amino acid (O, N, S) to which the acidic proton is bonded in the protonated state. The distributions are shown in Figure 3 for all amino acids investigated as well as for aqueous hydronium ions. Peak positions and coordination numbers obtained by integration of the
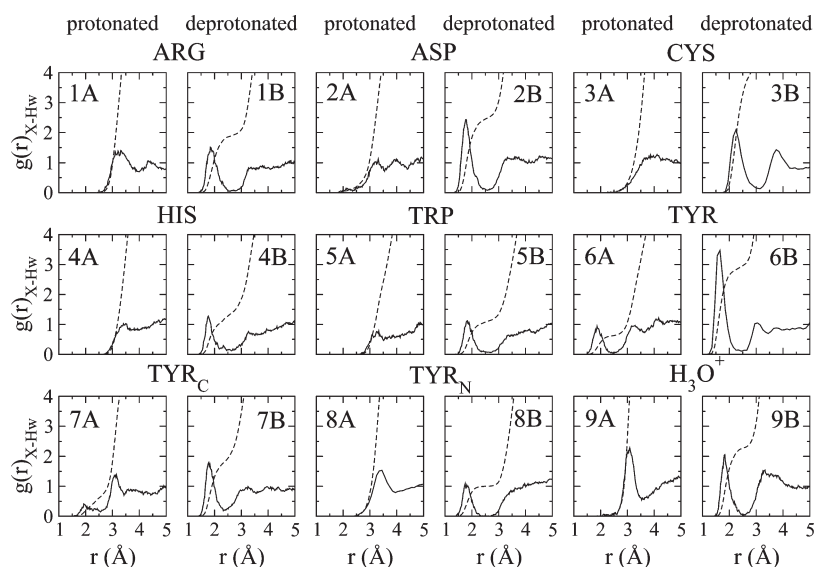
Figure 3. Radial distribution functions $g_{X-Hw}(r)$ (solid lines) and coordination numbers (dashed lines) for protonated and deprotonated amino acids. H_W denotes the hydrogen atoms of water molecules (solvent). X denotes the basic N(H) atom of the side chain of Arg_{mod} (1A, 1B), the O(H) atom of the side chain of Asp (2A, 2B), the S(H) atom of the side chain of Cys (3A, 3B), the εN atom of the side chain of His (4A, 4B), the N(H) atom of the side chain of Trp* (5A, 5B), the O atom of the side chain of Tyr (6A, 6B), the O(H) atom of the carboxyl group of Tyr_C (7A, 7B), the $N(H_3)$ atom of Tyr_N (8A, 8B), and the O atom of hydronium/water, respectively (9A, 9B). Coordination numbers are obtained by spherical integration of $g_{X-Hw}(r)$. The positions of the first maxima and coordination numbers are given in Table 5.

Table 5. Characterization of the First Solvation Shell of Protonated and Deprotonated Amino Acids, As Obtained from Density Functional Based Molecular Dynamics Simulation

	protonated		deprotonated		
	CN^a	$R_0{}^b$ (Å)	CN^a	$R_0{}^b$ (Å)	Δn^c
Arg_{mod}	0.0	3.3	1.7	1.8	1.7
Asp	0.1	3.3	2.4	1.7	2.3
Cys	0.3	4.0	4.1	2.3	3.8
His	0.0	3.4	1.2	1.8	1.2
Trp*	0.0	3.3	1.0	1.8	1.0
Tyr	0.6	1.9	2.9	1.6	2.3
Tyr_C	0.5	2.0	1.8	1.8	1.3
Tyr_N	0.0	3.4	1.0	1.8	1.0
H_3O^+	0.0	3.0	2.2	1.9	2.2

^{a} Coordination numbers obtained from spherical integration of the radial distribution functions g_{X-Hw}, shown in Figure 3, up to a distance of 2.4 Å (except for Cys integrated up to 3.0 Å). ^{b} Position of the first peak of g_{X-Hw} shown in Figure 3. ^{c} $\Delta n =$ CN(deprotonated) − CN(protonated).

radial distributions up to a distance of 2.4 Å or 3.0 Å for cysteine are summarized in Table 5.

Cysteine exhibits the most significant change in solvation structure upon removal of the acidic proton (see Figure 3 (3A and 3B)). In the protonated state, the side chain forms a hydrophobic cavity, as indicated by the large distance for the first peak at $R_0 = 4.0$ Å (Figure 3 (3A)). Removal of the HS proton and formation of a negative charge on the S atom causes a major reorganization of the solvation shell, which is indicated by the formation of four weak hydrogen bonds with water molecules at an S−H distance of 2.3 Å (see Figure 3 (3B)). Thus, the water

coordination is qualitatively similar to the one reported for aqueous hydroxide.[34,45,46] However, we observe that an additional hydrogen bond is formed with the ammonium group, giving in total an approximately quadratic pyramidal hydrogen bond network.

A strong reorganization of the hydrogen bond network can also be observed when tyrosine is deprotonated. The coordination number changes from 0.6 (Figure 3 (6A)) to 2.9 ((Figure 3 (6B)), concomitant with a significant increase in the strength of the hydrogen bonds, as indicated by the decrease in the peak position from $R_0 = 1.9$ to 1.6 Å. The three water molecules bound to the tyrosinate oxyanion form a tetrahedral coordination, implying that the oxygen atom is better described as sp^3 hybridized rather than sp^2. Considering the carboxyl group of tyrosine, we find that in the protonated state the oxygen atom OH forms occasionally one weak hydrogen bond and, in the deprotonated state, 1.8 hydrogen bonds. Interestingly, this is less than for aspartate, which forms 2.4 hydrogen bonds, and aqueous formate.[28] This can can be explained by the fact that the positively charged α-ammonium group of tyrosine is in close proximity to its carboxylate group, thereby replacing the stabilizing effect of an additional hydrogen bond donating water molecule. The second oxygen atom of the carboxylate group of tyrosine points in a direction opposite the ammonium group, with the result that the coordination number (= 2.3) is very similar to the one obtained for aspartate.

Turning to the amino acids Arg_{mod}, His, Trp*, and Tyr_N, we find that in the protonated state the positively charged nitrogen atom does not form any hydrogen bonds with the solvent. Upon deprotonation, the nitrogen atoms of His, Trp*, and Tyr_N form one strong hydrogen bond in accord with expectation. By contrast, for Arg_{mod}, we find a coordination number of 1.7. One hydrogen bond is due to a water molecule strongly bonded to the sp^2 lone pair of the imide nitrogen as well as to the carboxylate group. A second hydrogen bond is formed by water

molecules that penetrate the space perpendicular to the molecular plane and interact with the Π-bonding orbital of the imide. This hydrogen bond is weak and of a temporary nature, thus giving only a fractional contribution of 0.7.

5. DISCUSSION

Our computations reproduce experimental absolute pK_a values of natural amino acids rather well judging from the reasonably small UME of 2.1 pK_a units. A similar accuracy was reported in our previous study on smaller inorganic and organic aqueous acids.[31] However, the present investigation differs in one important aspect from our previous work.[30,31] In refs 30 and 31, the aqueous proton was modeled as a hydronium ion in its Eigenstructure, whereas in current work, we treat the proton as a solvated particle that is not attached to a particular water molecule (it is modeled as a hydronium ion only in the intermediate states of the thermodynamic cycle, see eqs 10−12). By assuming that the proton forms a stable hydronium ion in water, one overestimates its free energy by a contribution $\Delta A_{H_3O^+}$, corresponding to the free energy of the reverse process of reaction eq 10. The present approach takes into account $\Delta A_{H_3O^+}$, whereas our previous formulation did not.

An accurate estimation of $\Delta A_{H_3O^+}$ is difficult, however, due to the delocalized nature of the "true" solvated proton, and the limited length of the DFMD trajectories. Adopting a simple quantum-classical model for the hydronium ion, we found that at ambient temperature the free energy difference between the solvated proton and the solvated hydronium is approximately equal to the ideal translational free energy of the proton, see eq 35. The result is that the absolute pK_a values are now shifted down by 0.19 eV with respect to our previous approach.[30,31] Indeed, a similar down shift of pK_a values by 0.2 eV was obtained previously, by fitting the calculated deprotonation free energies of small acids to the experimental pK_a values.[31] This gives some credence to our assumption that the reaction free energy of reaction eq 10 is dominated by the standard translational free energy of the proton.

We have recalculated the pK_a values reported in ref 31 according to our revised scheme and combined them with the results obtained herein for the amino acids. This gives in total a set of 18 inorganic and organic acids and natural amino acids for which we obtain a UME of 2.3 pK_a units and a SME of 2.0 pK_a units. The relatively small difference between UME and SME implies that pK_a values are in general overestimated. Though, for the amino acids studied here, the difference between UME and SME (2.1 and 1.3, respectively) is larger, implying a less systematic trend in the errors for this subset of acids. This could be related to the statistical uncertainty of the computed pK_a values, which is on the same order of magnitude as the UME. A breakdown of the UME into groups with the same functional groups reflects the overall trend (hydroxyl, 1.9; carboxylic acid, 2.1; amine/imine, 2.4), except for the thiol group which has a larger UME of 3.4 pK_a units.

There are a number of sources for errors in our pK_a calculations in addition to the modeling of the solvated proton. This becomes evident when the pK_a value *differences* of amino acids are compared, for which the solvation free energy of the proton drops out. The UME and the SME of the corresponding relative pK_a values are 2.4 and 0.2 when averaged over all possible pairs of amino acids. Since the SME is close to zero, the deviations are evenly spread in both directions and therefore not systematic.

They are likely to be the result of a number of factors. First, the three point approximation of the thermodynamic integral eq 21 may introduce some error. In previous work we have shown that this approximation gives accurate results for aqueous HCl and H_2S, but this may not be the case in general.[31] Unfortunately, using a finer grid for the integration was computationally not feasible for the relatively large system sizes studied. Second, the amino acids were modeled in their zwitterionic state in accord with their experimental protonation state at the pH where the side chain acidic groups titrate. Thus, the amino acids carry a large and permanent dipole that interacts with the acidic group of the side chain. While the structures were initiated from the minimum energy configuration and were found to be stable on the 10−20 ps time scale of our simulation, it may be that the charged groups forming the molecular dipole undergo fluctuations on a longer time scale, which could lead to shifts of the pK_a value that are not accounted for in the present simulations. Third, the finite number of solvent molecules used may also have some effect. We have used a simulation box large enough to include the first solvation shell of the amino acid. This is sufficient to model the response of the solvent in the vicinity of the acidic group. The effect of higher solvation shells on the energetics of deprotonation is expected to cancel to a good approximation for iso-Coulombic half reactions, i.e., for reactions where reactant and product have the same net charge. In this respect, it is interesting to note that the UME for the subsets of iso-Coulombic reactions is not lower than for the overall UME that includes all amino acid pairs. This indicates that the error is likely to be due to other reasons. Finally, there is also an error due to the approximation of the exchange-correlation functional. While proton affinities are generally very well described by the BLYP functional that was used in DFMD-TI, the UME of this functional for the PA8 set of molecules and atoms is not negligible on the pK_a energy scale, 1.53 kcal/mol or 1.1 pK_a units.[47] We expect, however, that the DFT error is smaller for proton affinity differences, which is the relevant quantity in pK_a calculations.

A comparison between the present DFMD-TI approach and QM-continuum calculations may be of interest. In ref 17, the performances of a number of QM-continuum calculation protocols with different continuum models were investigated on neutral organic and inorganic acids. It was found that the direct approach was unsuitable for calculation of absolute pK_a values, although reasonable results with UME < 3 and maximum error < 5 pK_a units were possible for alcohols and phenols in combination with the COSMO-RS continuum model. Hybrid approaches performed better with best UMEs of about 2 pK_a units, but the calculations were very sensitive on the number of explicit water molecules included. The proton exchange scheme, which yields pK_a values relative to a reference compound, gave good overall results with UMEs of 1−2 pK_a units for alcohols, phenols, and carboxylic acids. Thus, the error reported here for the DFMD-TI approach is comparable to the error of the proton-exchange scheme. Yet, this comparison can only be of a qualitative nature, since different compounds have been investigated with the two methods.

Finally, a word on the solvation structures obtained in this study. They generally comply with expectations, with a few exceptions. The oxyanion of tyrosinate forms three strong hydrogen bonds in a tetrahedral arrangement, implying that the oxygen atom is best described as sp^3 hybridized. This contrasts the situation in the gas phase, where oxygen is described as sp^2 hybridized, as required for delocalization of

1959

dx.doi.org/10.1021/ct100715x |*J. Chem. Theory Comput.* 2011, 7, 1951–1961

the negative charge over the phenyl ring. Evidently, the polarizing effect of the solvent leads to charge localization and a change in hybridization. We have also shown that in addition to solute—solvent interactions, also intramolecular interactions are important for the stabilization of the deprotonated species. In the cysteine anion, an approximately quadratic pyramidal hydrogen bond network is observed that includes the ammonium group. Similarly, one carboxylate oxygen of tyrosine is stabilized by the positive charge of the nearby ammonium group, whereas in modified arginine a water molecule binding to the imine group is tightly held in place by the nearby carboxylate group.

6. CONCLUSION

In conclusion, we have shown that DFMD-TI can reproduce absolute experimental acidity constants of naturally occurring amino acids to a UME of 2.1 pK_a units with a maximum absolute deviation of 4.0 pK_a units. The errors for absolute and relative pK_a values of the amino acids do not appear to be systematic. This is probably a consequence of the relatively small number of compounds investigated (limited by the high computational cost of DFMD compared to QM/continuum approaches) and the statistical uncertainty of our simulations of 1.1 pK_a units. Extending the set of amino acids with the set of acids studied previously,[30,31] we find a more systematic correlation in the errors. The computed pK_a values are on average overestimated by about 2 pK_a units. The accuracy reported here is comparable with that reported recently for the best QM/continuum calculations on a series of alcohols, phenols, and carboxylic acids.[17]

Our calculation of absolute pK_a values relies on a computational model for the solvated proton. We have used the Eigen-structure of the hydronium ion as in our previous calculations, but now corrected by a free energy term that accounts for the formation of the Eigen-structure from the "real", delocalized solvated proton. This correction term was estimated to be the standard translational free energy of the proton. It would be interesting and challenging to further investigate the accuracy of this approximation by explicit calculation of the free energy difference between the delocalized proton and the Eigen-structure of the hydronium ion.

The approximation made for the solvated proton is not the only source of errors in our calculations, because the average error in pK_a differences, for which the solvated proton drops out, is just as large. The accuracy of our approach can in principle be systematically improved (i) by carrying out longer simulations to reduce the statistical uncertainty, (ii) by using more integration points for evaluation of the thermodynamic integral, and (iii) by using a larger number of solvent molecules to reduce possible system size effects. With the steady increase in computational power, we remain confident that a more accurate and precise estimation of pK_a values from DFMD should be possible in the near future.

■ ASSOCIATED CONTENT

ⓢ Supporting Information.   The derivation of eq 24 and the results of the numerical calculation of ΔA_{Ad} using eq 24. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: j.blumberger@ucl.ac.uk.

■ REFERENCES

(1) Silva, C.; Da Silva, E.; Nascimento, M. *J. Phys. Chem. A* **2000**, *104* (11), 2402.
(2) Liptak, M.; Shields, G. *J. Am. Chem. Soc.* **2001**, *123* (30), 7314.
(3) Chipman, D. *J. Phys. Chem. A* **2002**, *106* (32), 7413.
(4) Klicic, J.; Friesner, R.; Liu, S.-Y.; Guida, W. *J. Phys. Chem. A* **2002**, *106* (7), 1327.
(5) Pliego, J.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.
(6) Saracino, G.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373*, 411.
(7) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493.
(8) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A. *J. Phys. Chem. A* **2007**, *111*, 4422.
(9) Lu, H.; Chen, X.; Zhan, C.-G. *J. Phys. Chem. B* **2007**, *111* (35), 10599.
(10) Jia, Z.; Du, D.; Zhou, Z.; Zhang, A.; Hou, R. *Chem. Phys. Lett.* **2007**, *439* (4−6), 374.
(11) Sadley-Sosnowska, N. *Theor. Chem. Acc.* **2007**, *118*, 281.
(12) Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B. *J. Phys. Chem. B* **2008**, *112*, 16860.
(13) Trummal, A.; Rummel, A.; Lippmaa, E.; Burk, P.; Koppel, I. A. *J. Phys. Chem. A* **2009**, *113*, 6206.
(14) Casasnovas, R.; Frau, J.; Ortega-Castro, J.; Salva, A.; Donoso, J.; Munoz, F. *J. Mol. Spectrosc.* **2009**, *912*, 5.
(15) Khalili, F.; Henni, A.; East, A. L. L. *J. Mol. Spectrosc.* **2009**, *916*, 1.
(16) Ho, J. M.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 295.
(17) Ho, J. M.; Coote, M. L. *Theor. Chem. Acc.* **2010**, *125*, 3.
(18) Eckert, F.; Diedenhofer, M.; Klamt, A. *Mol. Phys.* **2010**, *108*, 229.
(19) Simonson, T.; Carlsson, J.; Case, D. *J. Am. Chem. Soc.* **2004**, *126* (13), 4167.
(20) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157.
(21) Click, T. H.; Kaminski, G. A. *J. Phys. Chem. B* **2009**, *113*, 7844.
(22) Li, H.; Hains, A.; Everts, J.; Robertson, A.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106* (13), 3486.
(23) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109* (37), 17715.
(24) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253.
(25) Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M. *J. Am. Chem. Soc.* **2002**, *124*, 6594.
(26) Ivanov, I.; Chen, B.; Raugei, S.; Klein, M. *J. Phys. Chem. B* **2006**, *110* (12), 6365.
(27) Simon, C.; Ciccotti, G.; Klein, M. *Chem. Phys. Chem.* **2007**, *8* (14), 2072.
(28) Maurer, P.; Iftimie, R. *J. Chem. Phys.* **2010**, *132*, 074112.
(29) Sulpizi, M.; Sprik, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5238.
(30) Cheng, J.; Sulpizi, M.; Sprik, M. *J. Chem. Phys.* **2009**, *131*, 154504.
(31) Sulpizi, M.; Sprik, M. *J. Phys.: Condens. Matter* **2010**, *22*, 284116.
(32) Cheng, J.; Sprik, M. *J. Chem. Theory Comput.* **2010**, *6*, 880.
(33) Costanzo, F.; Della Valle, R. G.; Sulpizi, M.; Sprik, M. Submitted.
(34) Marx, D. *Chem. Phys. Chem.* **2006**, *7*, 1848.
(35) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(36) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(37) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785.

(38) Quickstep, version 2.0.0, CP2K Developers Group. http://www.cp2k.berlios.de (accessed August 2008).

(39) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S. Kollman, P. A. *AMBER 9*, University of California: San Francisco, CA, 2006.

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

(41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

(42) Seidel, R.; Faubel, M.; Winter, B.; Blumberger, J. *J. Am. Chem. Soc.* **2009**, *131*, 16127.

(43) Moens, J.; Seidel, R.; Geerlings, P.; Faubel, M.; Winter, B.; Blumberger, J. *J. Phys. Chem. B* **2010**, *114*, 9173.

(44) Blumberger, J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5651.

(45) Blumberger, J.; Ensing, B.; Klein, M. L. *Angew. Chem., Int. Ed.* **2006**, *45*, 2893.

(46) Blumberger, J.; Klein, M. L. *Chem. Phys. Lett.* **2006**, *422*, 210.

(47) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(48) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 75th ed.; CRC Press: Boca Roton, FL, 1995.

(49) Solar, S.; Getoff, N.; Surdhar, P.; Armstrong, D.; Singh, A. *J. Phys. Chem.* **1991**, *95* (9), 3639.