



pK_a calculations for tautomerizable and conformationally flexible molecules: partition function vs. state transition approach

Nicolas Tielker¹ · Lukas Eberlein¹ · Christian Chodun¹ · Stefan Güssregen² · Stefan M. Kast¹

Received: 17 February 2019 / Accepted: 7 April 2019 / Published online: 30 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Calculations of acidities of molecules with multiple tautomeric and/or conformational states require adequate treatment of the relative energetics of accessible states accompanied by a statistical-mechanical formulation of their contribution to the macroscopic pK_a value. Here, we demonstrate rigorously the formal equivalence of two such approaches: a partition function treatment and statistics over transitions between molecular tautomeric and conformational states in the limit of a theory that does not require adjustment by empirical parameters correcting energetic values. However, for a frequently employed correction scheme, linear scaling of (free) energies and regression with respect to reference data taking an additive constant into account, this equivalence breaks down if more than one acid or base state is involved. The consequences of the resulting inconsistency are discussed on our datasets developed for aqueous pK_a predictions during the recent SAMPL6 challenge, where molecular state energetics were computed based on the “embedded cluster reference interaction site model” (EC-RISM). This method couples integral equation theory as a solvation model to quantum-chemical calculations and yielded a test set root mean square error of 1.1 pK units from a partition function ansatz. For all practical purposes, the present results indicate that a state transition approach yields comparable accuracy despite the formal theoretical inconsistency, and that an additive regression intercept, which is strictly constant in the limit of large compound mass only, is a valid approximation.

Keywords pK_a prediction · Solvation model · Quantum chemistry · Integral equation theory · EC-RISM · Tautomers/conformers

Introduction

Calculations of acidity constants, K_a , are important not only for practical purposes but also serve as important benchmarks for testing solvation models used in conjunction with quantum-chemical calculations, particularly in an aqueous en-

vironment [1]. Without loss of generality, and specializing to water as a solvent, the constitutive reaction equation



is characterized thermodynamically by the relation between equilibrium constant, activities a , standard Gibbs energy of reaction $\Delta_r G^0$ and standard chemical potentials μ^0 via

$$\begin{aligned} -\beta^{-1} \ln K_a &= -\beta^{-1} \ln \frac{a(\text{H}_{\text{aq}}^+) a(\text{A}_{\text{aq}}^-)}{a(\text{HA}_{\text{aq}})} = \Delta_r G^0 \\ &= \mu^0(\text{H}_{\text{aq}}^+) + \mu^0(\text{A}_{\text{aq}}^-) - \mu^0(\text{HA}_{\text{aq}}) \end{aligned} \quad (2)$$

where β is an inverse temperature. The standard chemical potentials in solution are commonly referenced to a standard state of 1 bar and a formal concentration of $c^0 = 1$ M at the specified temperature (hereinafter assumed to be 298.15 K) under the assumption of infinite dilution.

This paper belongs to the Topical Collection Tim Clark 70th Birthday Festschrift

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00894-019-4033-4>) contains supplementary material, which is available to authorized users.

✉ Stefan M. Kast
stefan.kast@tu-dortmund.de

¹ Physikalische Chemie III, Technische Universität Dortmund, Otto-Hahn-Str. 4a, 44227 Dortmund, Germany

² R&D Integrated Drug Discovery, Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main, Germany

Quantum calculations of these quantities, for instance by employing a continuum solvation approach, usually model such an ideal solution state by construction, approximating the standard chemical potential of a compound i in a given, fixed conformational and tautomeric state j , for instance, by [2–5].

$$\mu_j^0(i) \approx \mu_j^{0,\text{id}}(i) + E_j^{0,\text{sol}}(i) + \mu_j^{0,\text{ex}}(i) + G_j^{0,\text{RRHO}}(i) \equiv \mu_j^{0,\text{id}}(i) + G_j^0(i). \quad (3)$$

It is given by the sum of an ideal (“id”) part, which contains the explicit reference to the standard concentration to be specified below, and an interaction component, termed $G_j^0(i)$ here. The latter can be approximated by adding an electronic energy in solution, $E_j^{0,\text{sol}}(i)$, an excess chemical potential, $\mu_j^{0,\text{ex}}(i)$, that represents the Gibbs energy of solvation upon transferring a solute in the “frozen” structural and electronic solution state from the (ideal) gas phase into the solvent (assuming identical formal gas and solution phase concentrations), and potentially a “rigid rotor, harmonic oscillator” (RRHO) model of rotational and vibrational contributions to the Gibbs energy. As the standard condition of infinite dilution is implicitly assumed by constructing the Hamiltonian, the superscript “0” at the interaction terms $G_j^0(i)$ can be dropped for simplicity.

For treating multistate species comprising an ensemble of distinct tautomeric and conformational states, two strategies are available. One approach is to sum over states by defining, in a more or less ad hoc manner, a canonical partition function while ignoring pressure-volume contributions [3–5], to end up for the protonation equilibrium with

$$\Delta_r G^0 = \mu^{0,\text{id}}(\text{H}_{\text{aq}}^+) + \mu^{0,\text{id}}(\text{A}_{\text{aq}}^-) - \mu^{0,\text{id}}(\text{HA}_{\text{aq}}) + \mu^{\text{ex}}(\text{H}_{\text{aq}}^+) - \beta^{-1} \ln \frac{\sum_{j=1}^M \exp[-\beta G_j(\text{A}_{\text{aq}}^-)]}{\sum_{k=1}^N \exp[-\beta G_k(\text{HA}_{\text{aq}})]} \quad (4)$$

where we sum over M base and N acid states. The fourth term on the right hand side (r.h.s.) represents the Gibbs energy of hydration of the “proton” (again assuming identical gas phase and solution state concentrations) [6, 7] and otherwise only ideal terms that are usually assumed to be an additive constant. Therefore, on the decadic pK scale we finally obtain the expression for the partition function (PF) approach,

$$pK_{\text{a}}^{\text{PF}} = \frac{\beta \Delta_r G^0}{\ln 10} = b - \frac{m}{\ln 10} \ln \frac{\sum_{j=1}^M \exp[-\beta G_j(\text{A}_{\text{aq}}^-)]}{\sum_{k=1}^N \exp[-\beta G_k(\text{HA}_{\text{aq}})]} \quad (5)$$

with

$$b = \frac{\beta}{\ln 10} [\mu^{0,\text{id}}(\text{H}_{\text{aq}}^+) + \mu^{0,\text{id}}(\text{A}_{\text{aq}}^-) - \mu^{0,\text{id}}(\text{HA}_{\text{aq}}) + \mu^{\text{ex}}(\text{H}_{\text{aq}}^+)], \quad (6)$$

where the terms assumed to be constant in total (i.e., ideal gas and proton) are contained in b and, additionally, computational flexibility is offered by introducing a parameter m , which, ideally, is 1. The parameters m and b are typically adjusted by fitting to experimental reference data, as was done by us [3, 4, 8] and others [5, 9, 10] (the latter reference also representing an early example of a PF-type treatment), to name just a few.

The alternative is to connect all base and acid states by individual transition equilibria as



for which straightforward reduction of Eq. (5) would give

$$pK_{\text{a},jk} = b + \frac{m\beta}{\ln 10} [G_j(\text{A}_{\text{aq}}^-) - G_k(\text{HA}_{\text{aq}})]. \quad (8)$$

The individual state-to-state equilibrium constants can then be assembled to yield the macroscopic form from mass balance,

$$K_{\text{a}} = \frac{a(\text{H}_{\text{aq}}^+) \sum_{j=1}^M c_j(\text{A}_{\text{aq}}^-) / c^0}{\sum_{k=1}^N c_k(\text{HA}_{\text{aq}}) / c^0}, \quad (9)$$

as [11].

$$K_{\text{a}}^{\text{ST}} = \sum_{j=1}^M \frac{1}{\sum_{k=1}^N \frac{1}{K_{\text{a},jk}}} \quad (10)$$

where “ST” denotes the state transition approach. Here, in contrast to the analysis by Bochevarov et al. [11], who distinguish between “micro-” (i.e., tautomeric) and “nano-” (i.e., conformational) states, which leads to another layer in the continued fraction expansion, we need no such discrimination as the concepts of tautomers and (underlying) conformers is purely semantic, though pragmatically useful for certain models [11]. Physically, tautomers and conformers for a certain ionization state refer to different local minima of the (free) energy surface derived from one and the same molecular Hamiltonian, whereas the Hamiltonians of acid and base forms differ. Hence, we simply refer to “states” between which transitions can occur in the equilibrium mixture.

Though plausible, as the information content of both the PF and the ST approaches in terms of the state-specific Gibbs energies is identical, it is not immediately obvious under which circumstances both methods yield identical results. The goal of the present work was therefore to elucidate the formal equivalence both analytically and numerically. As will be shown below, both methods agree only in the limiting cases of a regression “slope” parameter m being exactly 1, i.e., for ideal (and usually inapplicable) models that do not require any form of empirical scaling of energies, or, trivially, in situation where only single acid and base states are considered. Numerically, this conclusion will be illustrated and discussed by re-analysis of our previous results

obtained during the recent SAMPL6 (“Statistical Assessment of the Modeling of Proteins and Ligands”) challenge [4, 12] on blindly predicting aqueous pK_a values for a number of kinase inhibitor fragments with multiple protonation states and considerable conformational flexibility.

Theory

Formal correspondence of the PF and ST approaches can be proved if the mass balance equation leading to the continued fraction representation (10) can be derived on the same statistical-mechanical footing as Eq. (5) and its reduction to Eq. (8). We therefore start with the fundamental expression for the chemical potential of a molecule (omitting index i for notational simplicity) composed of distinct states j such that the approximation (3) and the assumption of negligible pressure-volume work hold,

$$\begin{aligned} \mu &= -\beta^{-1} \ln \frac{V}{\Lambda^3 N_M} - \beta^{-1} \ln \sum_j \exp[-\beta G_j] \\ &= \mu^{\text{id}} - \beta^{-1} \ln Z \equiv \mu^{\text{id}} + G. \end{aligned} \tag{11}$$

Here, G is the excess (interaction) part of the total chemical potential as in Eq. (3), V represents the volume, N_M is the number of solute molecules which is 1 at infinite dilution, Z is the partition function, and Λ denotes the thermal wavelength given by

$$\Lambda = \left(\frac{\beta h^2}{2\pi m_M} \right)^{1/2} \tag{12}$$

with molecular mass m_M and Planck’s constant h . The statistical-mechanical chemical potential should be equivalent to the macroscopic thermodynamic definition

$$\begin{aligned} \mu &= \mu^0 + \beta^{-1} \ln a = \mu^0 + \beta^{-1} \ln(\gamma c/c^0) \\ &= \mu^0 + \beta^{-1} \ln \sum_j \gamma_j c_j / c^0 \stackrel{c \rightarrow 0}{=} \mu^0 + \beta^{-1} \ln \sum_j c_j / c^0 \end{aligned} \tag{13}$$

with total solute concentration c , split into state contributions c_j according to mass balance, and activity coefficients γ_j that approach 1 at infinite dilution. By noting that the probability of a state j can be written as

$$p_j = \frac{\exp[-\beta G_j]}{Z} = \frac{c_j}{c} = \frac{c_j}{c} \frac{c^0}{c^0} \tag{14}$$

where we inserted $1 = c^0/c^0$, and inserting $1 = \sum_j p_j$ as denominator in G of Eq. (11) we have

$$\begin{aligned} G &= -\beta^{-1} \ln \frac{\sum_j \exp[-\beta G_j]}{\sum_j p_j} = +\beta^{-1} \ln \sum_j \frac{p_j}{Z} = +\beta^{-1} \ln \left(\frac{c^0}{Zc} \sum_j \frac{c_j}{c^0} \right) \\ &= \beta^{-1} \ln \frac{c^0}{Zc} + \beta^{-1} \ln \sum_j \frac{c_j}{c^0}. \end{aligned} \tag{15}$$

We recover the concentration dependence of (13) as the last term on the r.h.s., and the standard chemical potential therefore becomes

$$\mu^0 = \mu^{\text{id}} + \beta^{-1} \ln \frac{c^0}{Zc} = -\beta^{-1} \ln \frac{ZVc}{c^0 \Lambda^3} \tag{16}$$

which, by also noting that $c \rightarrow 1/V$ at infinite dilution, finally yields

$$\mu^0 = -\beta^{-1} \ln \frac{Z}{c^0 \Lambda^3} = -\beta^{-1} \ln \frac{\sum_j \exp[-\beta G_j]}{c^0 \Lambda^3}. \tag{17}$$

For the protonation equilibrium in the partition function derivation we then ultimately find from inserting the expressions for the standard chemical potential for the reacting species into Eq. (2) and taking the negative decadic logarithm

$$\begin{aligned} pK_a &= -\frac{1}{\ln 10} \ln \frac{\Lambda^3(\text{HA})(c^0)^{-1}}{\Lambda^3(\text{A}^-)\Lambda^3(\text{H}^+)} \\ &\quad + \frac{1}{\ln 10} \mu^{\text{ex}}(\text{H}^+_{\text{aq}}) - \frac{1}{\ln 10} \ln \frac{\sum_{j=1}^M \exp[-\beta G_j(\text{A}^-_{\text{aq}})]}{\sum_{k=1}^N \exp[-\beta G_k(\text{HA}_{\text{aq}})]}. \end{aligned} \tag{18}$$

Comparison with Eq. (5) shows that both relations are equivalent (for $m = 1$), showing that mass balance leads directly to the partition function approach. It is, however, important to note that the regression intercept b , i.e., the sum of the first two terms in the latter equation is actually *not* a constant as it depends not only on the mass of proton but also on the mass ratio of acid and base forms via the thermal wavelengths, though not on the particular state. Unless the mass-dependent terms are grouped with the Boltzmann factors the intercept can be interpreted as essentially constant only in the limit of much larger molecular mass of the compound compared to the proton, which, however, holds true in most situations. In this limit, the first term becomes $-4.39 \text{ kcal mol}^{-1}$ (see also [13]) compared to the much larger Tissandier value for the Gibbs solvation energy of the proton of $-265.89 \text{ kcal mol}^{-1}$ [6] (assuming identical gas and solution phase concentrations). For, e.g., HF, the first quantity would change by 7.5%, corresponding to ca. 0.24 pK units. Very accurate calculations should, therefore, take this effect into account. Note that this result holds not only within the quantum-statistical formalism invoked for the chemical potential, but also in a classical framework, where integration over momentum space yields the identical dependence of the standard chemical potential on molecular mass and standard concentration (see Eq. (8) in [14]).

To close the proof of equivalence, mass balance also leads to the continued fraction expansion (10) where Eq. (8) can be inserted to show under which conditions Eqs. (5) and (18) arise. Mass balance according to Eq. (9) for the protonation

equilibrium readily leads to the continued fraction expansion (10) as derived in [11]. Rewriting Eq. (8) on the energy scale and inserting into (10) yields

$$K_a^{ST} = \sum_{j=1}^M \frac{1}{\sum_{k=1}^N \frac{1}{K_{a,jk}}} \\ = \sum_{j=1}^M \frac{1}{\sum_{k=1}^N \left(\frac{\exp[-\beta G_k(\text{AH}_{\text{aq}})]}{\exp[-\beta G_j(\text{A}_{\text{aq}}^-)]} \right)^m} 10^b. \quad (19)$$

In the innermost sum, the j -dependent denominator is constant for all k such that we obtain

$$K_a^{ST} = \sum_{j=1}^M \frac{1}{\frac{\sum_{k=1}^N \exp[-\beta G_k(\text{AH}_{\text{aq}})]^m}{\exp[-\beta G_j(\text{A}_{\text{aq}}^-)]^m} 10^b} = \sum_{j=1}^M \frac{\exp[-\beta G_j(\text{A}_{\text{aq}}^-)]^m 10^{-b}}{\sum_{k=1}^N \exp[-\beta G_k(\text{AH}_{\text{aq}})]^m} \\ = 10^{-b} \frac{\sum_{j=1}^M \exp[-\beta G_j(\text{A}_{\text{aq}}^-)]^m}{\sum_{k=1}^N \exp[-\beta G_k(\text{AH}_{\text{aq}})]^m} \quad (20)$$

in the ST form. In contrast, the corresponding PF result derived from Eq. (5) reads

$$K_a^{\text{PF}} = 10^{-b} \left(\frac{\sum_{j=1}^M \exp[-\beta G_j(\text{A}_{\text{aq}}^-)]}{\sum_{k=1}^N \exp[-\beta G_k(\text{HA}_{\text{aq}})]} \right)^m \quad (21)$$

which clearly shows that both expressions can only be identical if either $m = 1$ for multistate mixtures or if only one state per acid and base form exists while b is unaffected.

Numerical illustration

To demonstrate the effect of slope parameters $m \neq 1$ on the relative performance of both the PF and ST models for a realistic prediction problem, here we re-analyze training and test set data obtained during the SAMPL6 challenge [4], where the PF model was employed exclusively. Briefly, we tested EC-RISM [2] theory for treating aqueous solvation in conjunction with quantum-chemical calculations, and were able to show that root mean square errors (RMSE) of ca. 1.0 pK units could be achieved for a well-known training set [15]. About the same error (1.1 pK units) was obtained for the independent test set composed of kinase inhibitor fragments, whose microstates were provided as part of the SAMPL6 challenge where the task was to blindly predict their pK_a values. The challenge was explicitly designed to cover molecules with multiple protonation sites, ionization states, and high conformational freedom, which necessitated adequate conformational sampling based on a large reference set of

tautomers provided by the organizers. It is therefore not immediately clear that the PF and the ST approaches should perform similarly, as a non-unity slope together with large state ensembles suggests discrepancies (see derivation and discussion above).

We confine our re-analysis to the best-performing quantum-chemical level of theory and solvation model, termed “MP2/6-311+G(d,p)/ $\varphi_{\text{opt}}/c_{\text{opt}2}$ ” in [4], where we used the two best-ranked conformations per tautomeric state (“ $c_{\text{opt}2}$ ”) with an optimized model to compute electrostatic solute-solvent interactions (“ φ_{opt} ”) in combination with the 6-311+G(d,p) basis set within MP2 calculations. As a consistency check, besides the “2par” regression models (m and b variable), we additionally tested the PF and the ST models with a fixed slope of $m = 1$ (“1par”), not only to demonstrate the resulting equivalence, but also to analyze the impact on predictive performance. All statistical regression and metrics data are found in Tables 1 (training set) and 2 (test set), while the individual correlations of calculated and experimental data for the various methods are depicted in Fig. 1. Note that, unlike the linear regression problem of the PF approach, the ST model requires nonlinear optimization of a loss function defined by the sum of squared residuals.

Training both models in 1par and 2par variants showed the expected results. While the trained parameters and the resulting statistical metrics are identical in the 1par case, as mathematically required, there is a small but almost

Table 1 Parameters of optimized embedded cluster reference interaction site model (EC-RISM-) based aqueous pK_a models for the training set along with statistical metrics [root mean square error (RMSE), mean absolute error (MAE), mean signed error (MSE), slope m , intercept b , and coefficient of determination R^2 from predictive regression]. “PF/2par” represents metrics reported in [4]

	RMSE	MAE	MSE	R^2	m	b
PF/2par						
All	1.04	0.87	0.00	0.93	0.74	-150.72
Acids	0.93	0.77	0.56	0.93		
Bases	1.14	0.97	-0.54	0.95		
ST/2par						
All	1.05	0.87	0.00	0.93	0.74	-150.64
Acids	0.93	0.77	0.56	0.93		
Bases	1.14	0.97	-0.54	0.95		
PF/1par						
All	1.66	1.43	0.00	0.93	1.00	-204.24
Acids	1.20	1.06	0.53	0.93		
Bases	2.00	1.78	-0.51	0.95		
ST/1par						
All	1.66	1.43	0.00	0.93	1.00	-204.24
Acids	1.20	1.06	0.53	0.93		
Bases	2.00	1.78	-0.51	0.95		

Table 2 Statistical metrics for pK_a predictions on the test set (RMSE, MAE, MSE, slope m' , intercept b' , and coefficient of determination R^2 from descriptive regression) for various models. “PF/2par” represents metrics reported in [4]

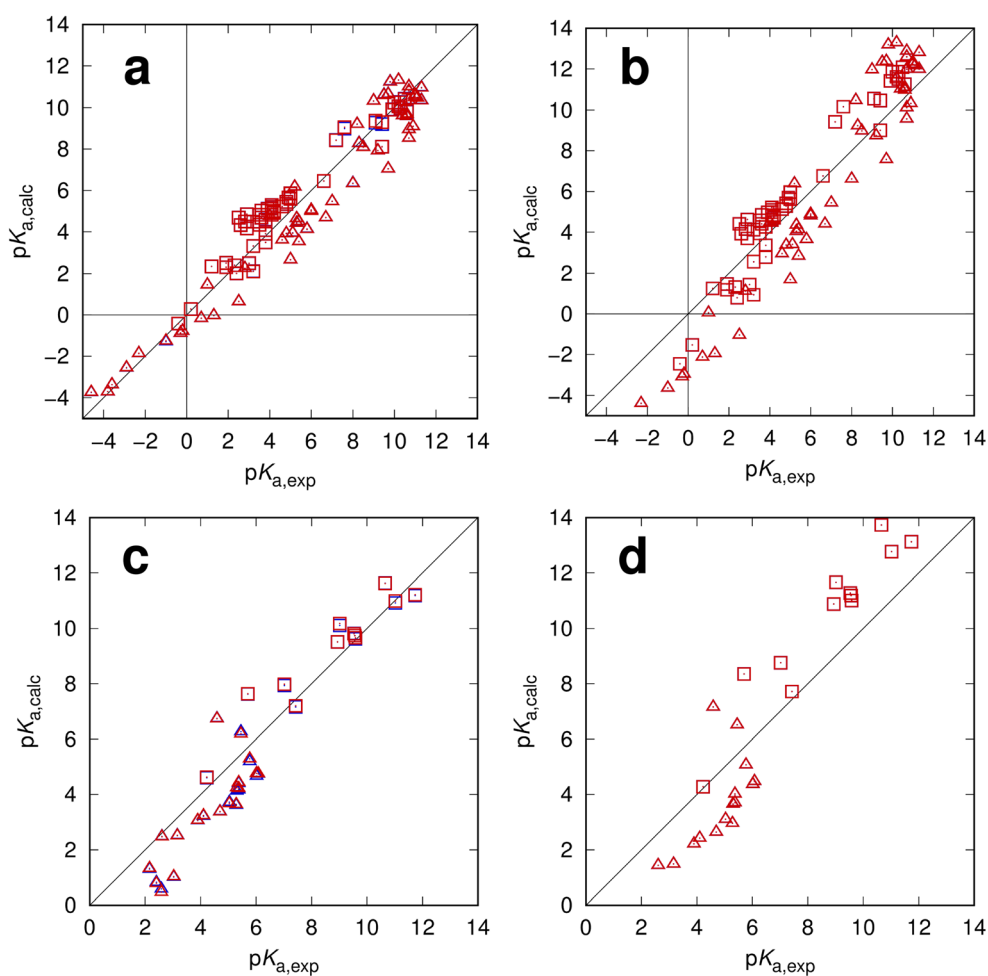
Model	RMSE	MAE	MSE	m'	b'	R^2
PF/2par						
All	1.13	0.97	-0.36	1.17	-1.38	0.91
Anionic	0.80	0.61	0.46	0.86	1.70	0.91
Cationic	1.30	1.20	-0.88	1.08	-1.25	0.69
ST/2par						
All	1.13	0.96	-0.35	1.17	-1.40	0.91
Anionic	0.81	0.61	0.48	0.86	1.72	0.91
Cationic	1.30	1.18	-0.88	1.10	-1.33	0.70
PF/1par						
All	2.04	1.87	-0.32	1.57	-3.75	0.91
Anionic	1.90	1.70	1.70	1.15	0.38	0.91
Cationic	2.12	1.97	-1.59	1.46	-3.58	0.68
ST/1par						
All	2.04	1.87	-0.32	1.57	-3.75	0.91
Anionic	1.90	1.70	1.70	1.15	0.38	0.91
Cationic	2.12	1.97	-1.59	1.46	-3.58	0.68

negligible difference in the results for the 2par models, which is mainly a result of the limited amount of tautomeric and conformational freedom in the training data set (see Online Resources in [4]). The same holds true when applying the trained models to the test dataset from the SAMPL6 challenge. Despite the drastically differing diversity, the results are in line with the results from the training set. One has to keep in mind, though, that the 1par models are substantially inferior regarding performance (training set), and even more so in terms of predictivity (test set), which emphasizes the importance of scaling (free) energies by the slope parameter. Surprisingly, since the differences between PF and ST models are so small, it is in practice almost irrelevant which approach is preferred for acidity predictions.

Concluding remarks

In summary, we addressed a conceptual problem for practical pK_a calculations that results from the necessity to include an energy scaling parameter m into the prediction model that is

Fig. 1a–d Embedded cluster reference interaction site model (EC-RISM-)derived vs. experimental pK_a for training [15] (top row) and test set [4] (bottom row) calculated with either the partition function (PF; blue) or the state transition (ST) approach (red, mostly overlaying blue symbols). The charge state of the molecular species involved in the transition is reflected by different symbols: squares acids (training set)/anionic transition (test set), triangles bases (training set)/cationic transition (test set). Anionic and cationic refer to the sum over the charges of all species involved in a reaction which is negative or positive, respectively. Pairs of calculated and experimental data points are collected in Online Resource 1, raw data underlying the regression analysis was taken from Online Resources to [4]. **a, c** using m as a free parameter and **b, d**: using a fixed $m = 1$.



typically adjusted empirically. To this end, we derived the rigorous statistical-mechanical expressions for the acidity constants for two variants of multistate calculations, which revealed the source of an inconsistency when used within regression analysis.

From a mathematical perspective, the issue boils down to the inequality $(x + y)^m \neq x^m + y^m$ for arbitrary $m \neq 1$ where x and y represent non-zero Boltzmann factors of different tautomeric or conformational states of a given molecule. This finding is an example of a case in which *formal* equivalence of two approaches does not necessarily translate into equivalence in practical applications where numerical model adjustments turn out to be necessary. One might have expected significant differences between the partition function (l.h.s.) and the state transition (r.h.s.) approaches as the regression results indicate significant deviations from 1. However, for the training set the largest difference between PF and ST results is on the order of 0.1 pK units with our slope parameter of 0.74. Even with a smaller m of 0.5, this difference would not exceed approximately 0.2 pK units. This means that both models are quantitatively very similar for practical purposes, at least as long as a sufficiently accurate methodology is applied as in this work, and there is no obvious reason to prefer one method over the other.

Another result of the rigorous derivation was that the regression constant is actually variable, though with limited, and, in practice, mostly negligible range, as it depends, strictly speaking, on the mass ratio between acid and base form, which approaches unity only in the limit of large molecules. Taken together, these findings could be useful to the community as they clarify potential sources of controversy.

Acknowledgments This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2033 – Projektnummer 390677874, and under the Research Unit FOR 1979. We also thank the IT and Media Center (ITMC) of the TU Dortmund for computational support and, of course, Tim Clark for the continuous fruitful collaborations and discussions over the years.

References

- Alongi KS, Shields GC (2010) Theoretical calculations of acid dissociation constants: a review article. *Ann Rep Comput Chem* 6:113–138
- Kloss T, Heil J, Kast SM (2008) Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach. *J Phys Chem B* 112:4337–4343
- Tielker N, Tomazic D, Heil J, Kloss T, Ehrhart S, Güssregen S, Schmidt KF, Kast SM (2016) The SAMPL5 challenge for embedded-cluster integral equation theory: solvation free energies, aqueous pK_a , and cyclohexane–water log D . *J Comput Aided Mol Des* 30:1035–1044
- Tielker N, Eberlein L, Güssregen S, Kast SM (2018) The SAMPL6 challenge on predicting aqueous pK_a values from EC-RISM theory. *J Comput Aided Mol Des* 32:1151–1163
- Pracht P, Wilcken R, Udvarhelyi A, Rodde S, Grimme S (2018) High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pK_a values in the context of the SAMPL6 challenge. *J Comput Aided Mol Des* 32:1139–1149
- Tissandier MD, Cowen KA, Feng AY, Gundlach E, Cohen MH, Earhart AD, Coe JV (1998) The Proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *J Phys Chem A* 102:7787–7794
- Zhang H, Jiang Y, Yan H, Cui Z, Chunhua Y (2017) Comparative assessment of computational methods for free energy calculations of ionic hydration. *J Chem Inf Model* 57:2763–2775
- Heil J, Tomazic D, Egbers S, Kast SM (2014) Acidity in DMSO from the embedded cluster integral equation quantum solvation model. *J Mol Model* 20:2161
- Klamt A, Eckert F, Diedenhofen M, Beck ME (2003) First principles calculations of aqueous pK_a values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the pK_a scale. *J Phys Chem A* 107:9380–9386
- Beck ME, Bürger T (2003) Predicting acidity for agrochemicals. In: Ford M, Livingstone D, Dearden J, Van deWaterbeemd H (eds) *Euro-QSAR 2002: designing drugs and crop protectants*. Blackwell, Oxford, pp 446–450
- Bochevarov AD, Watson MA, Greenwood JR (2016) Multiconformation, density functional theory-based pK_a prediction in application to large, flexible organic molecules with diverse functional groups. *J Chem Theory Comput* 12:6001–6019
- <https://drugdesigndata.org/about/sampl6>. Accessed 13 February 2019; see also special issue of *J Comput Aided Mol Design* (2018) 32(10)
- Rebollar-Zepeda A, Galano A (2016) Quantum mechanical based approaches for predicting pK_a values of carboxylic acids: evaluating the performance of different strategies. *RSC Adv* 6:112057
- Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* 72:1047–1069
- Klicic JJ, Friesner RA, Liu SY, Guida WC (2002) Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. *J Phys Chem A* 106:1327–1335

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.