J Mol Model (2019) 25:48
https://doi.org/10.1007/s00894-019-3928-4

ORIGINAL PAPER

Correlation between molecular acidity (pK_a) and vibrational spectroscopy

Niraj Verma[1] · Yunwen Tao[2] · Bruna Luana Marcial[3] · Elfi Kraka[1]

Received: 20 November 2018 / Accepted: 3 January 2019

Abstract

Molecular acidity is an important physicochemical property, which is often represented by the pK_a value as the measure of acidity strength. However, the accurate calculation and prediction of pK_a values is still an unsolved problem for computational chemistry. In this work, we present for the first time a direct correlation between pK_a values and local vibrational frequencies for 15 different groups of compounds with various substituents. This correlation was derived from a quadratic function of two selected local vibrational frequencies as independent variables used to characterize electronic structure features influencing the molecular acidity. In total, 180 molecules were investigated with this correlation model. For each group of molecules, we found a strong correlation with root mean squared errors and mean absolute errors of less than 0.11 and 0.09 pK_a units, respectively. The correlation between pK_a and local vibrational modes, established in this work, can be generally applied to all compounds whose pK_a values are dominated by electronic substituent effects. In this regard, the new correlation model constitutes a powerful link between the well-known Hammett equation and vibrational spectroscopy. Furthermore, it allows a quick prediction of the pK_a values for new group members with different substituents.

Keywords Vibrational spectroscopy · pK_a value · Linear regression · Local vibrational mode analysis · Hammett equation

Niraj Verma and Yunwen Tao contributed equally to this work.

This paper belongs to Topical Collection QUITEL 2018 (44th Congress of Theoretical Chemists of Latin Expression)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00894-019-3928-4) contains supplementary material, which is available to authorized users.

✉ Elfi Kraka
ekraka@gmail.com

Niraj Verma
nirajverma288@gmail.com

Yunwen Tao
ywtao.smu@gmail.com

Bruna Luana Marcial
bruna.marcial@ifgoiano.edu.br

[1] Computational and Theoretical group (CATCO), Department of Chemistry, Southern Methodist University, 3215 Daniel Ave, Dallas, TX 75275-0314, USA

[2] Department of Chemistry, New York University, 100 Wash. Sq. East, New York, NY 10003, USA

[3] Rodovia BR153, KM633 - Zona Rural, Morrinhos, GO 75650-000, Brazil

Introduction

Molecular acidity, the ability or tendency of an acid to lose a proton, is one of the most fundamental and important physiochemical properties of a molecule, which is essential in many chemical and biological processes [1–17]. Upon dissociation, an acid releases a proton, making the solution acidic. This processes is monitored by the equilibrium constant K_a of the corresponding dissociation reaction. For an acid HA, which dissociates into A^- and H^+, the negative logarithm of K_a is the pK_a of the corresponding acid with $K_a = [H^+][A^-]/[HA]$. For a charged acid BH^+ that dissociates to B and H^+, the pK_a is the negative logarithm of $K_a = [B][H^+]/[BH^+]$.

Despite the numerous attempts reported in the literature [5, 18–36], the accurate prediction of pK_a values by computational means is still an unsolved problem. A common approach is to use a thermodynamic cycle to calculate the pK_a by utilizing the standard Gibbs free energy change $\Delta G = 2.303\ RT\ pK_a$ [24, 37], applying wave function or density functional theory (DFT) methodologies [18, 38]. This procedure may suffer from errors caused by (i) the setup of the thermodynamic cycle and (ii) the way solvent and proton hydration is treated [39, 40]. Continuum

Published online: 30 January 2019

solvation models based on the quantum mechanical charge density of a solute molecule interacting with a continuum description of the solvent [41], introduced in 2009, are frequently applied for the calculation of pK_a values. Several modifications to this approach were made [42–51], allowing to estimate pK_a values in same cases within an accuracy of ca. 1 pK_a unit. An alternative approach is based on ab initio molecular dynamics (AIMD) simulations [52]. AIMD-based pK_a calculations require the simulation of the dissociation process. The challenge is to run the simulation long enough to capture the final state of the dissociation to be used for the prediction of the pK_a value. Correlating experimentally determined pK_a values with calculated molecular properties has also attracted attention [53–61], as it has been successful in special cases leading to a mean absolute error (MAE) of less than 0.5 pK_a units [1, 62].

The pK_a value is influenced by several factors such as solvent, configuration of the molecule, substituents, etc. Pioneering work by Oszczapowicz and coworkers [62] verified a Hammett-type correlation for a series of substituted amidines, which was extended to substituted benzoic acid by Huang, Liu, and coworkers [63] and to the benzoxaborole pharmacophore by Benkovic and coworkers [64], just to name a few examples. Overall, substituent effects play an important role [65], as they change the electronic structure of a molecule, even if the substitution occurs at a distance from the dissociating proton. Changing the electronic structure of a molecule via a substituent, the stability of the conjugate base is changed, and in this way the pK_a value. We have shown in recent work [66, 67] that the local vibrational modes, introduced by Konkoli and Cremer [68], sensitively reflect all electronic structure changes in a molecule upon substitution. Therefore, we expect a correlation between the local vibrational frequencies of the conjugate base and the pK_a value of a compound as sketched in Fig. 1.

The main objective of this work was to investigate whether the electronic structure changes resulting from different substituents are captured by some representative local vibrational frequencies of the conjugate base, and if so, a correlation between pK_a values and local vibrational frequencies exists. We tested this hypothesis for 15 different groups of molecules with different substituents R, (180 molecules in total), shown in Fig. 2. We selected five different pairs of representative local vibrational modes ω_1 and ω_2 illustrated in Fig. 3 and as described below. For each group, the local vibrational frequencies ω_1 and ω_2 were quadratically correlated to experimental pK_a values.

The correlation model applied in this work is defined in Eq. 1

$$pK_a = c1\times\omega_1 + c2\times\omega_2 + c3\times\omega_1^2 + c4\times\omega_2^2 + c5\times\omega_1\omega_2 + c6 \quad (1)$$

where c_i, $i \in \{1, ..., 5, 6\}$ are the correlation constants for each group, which were linearly optimized with regard to experimentally known pK_a values. For any new group member with a different substituent R, the corresponding pK_a can then be calculated via Eq. 1 as long as ω_1 and ω_2 are available.

The paper is arranged as follows. In Sect. 2 we describe the computational methods used for the calculation of the representative local mode frequencies ω_1 and ω_2 including a short description of the local vibrational mode analysis. In Sect. 3 we discuss the results for a total of 180

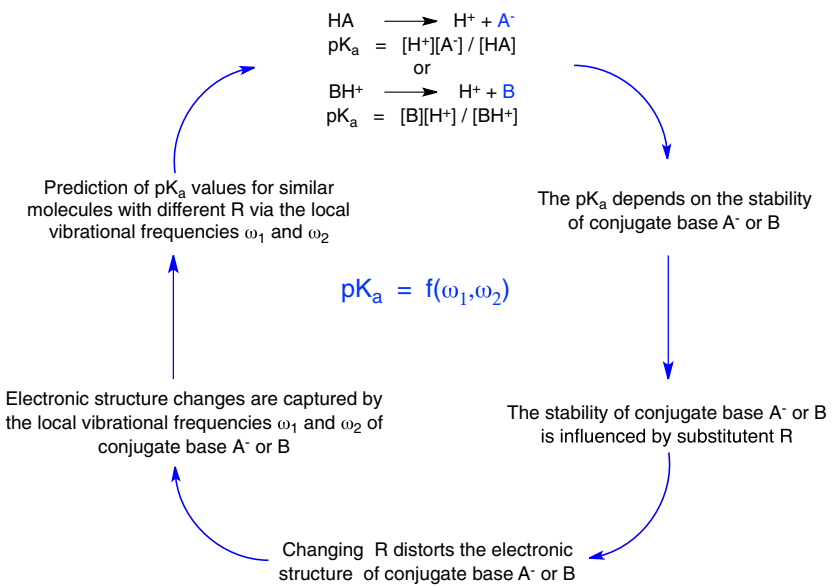

Fig. 1 Relationship between pK_a values and local vibrational frequencies for a conjugate base of an acid

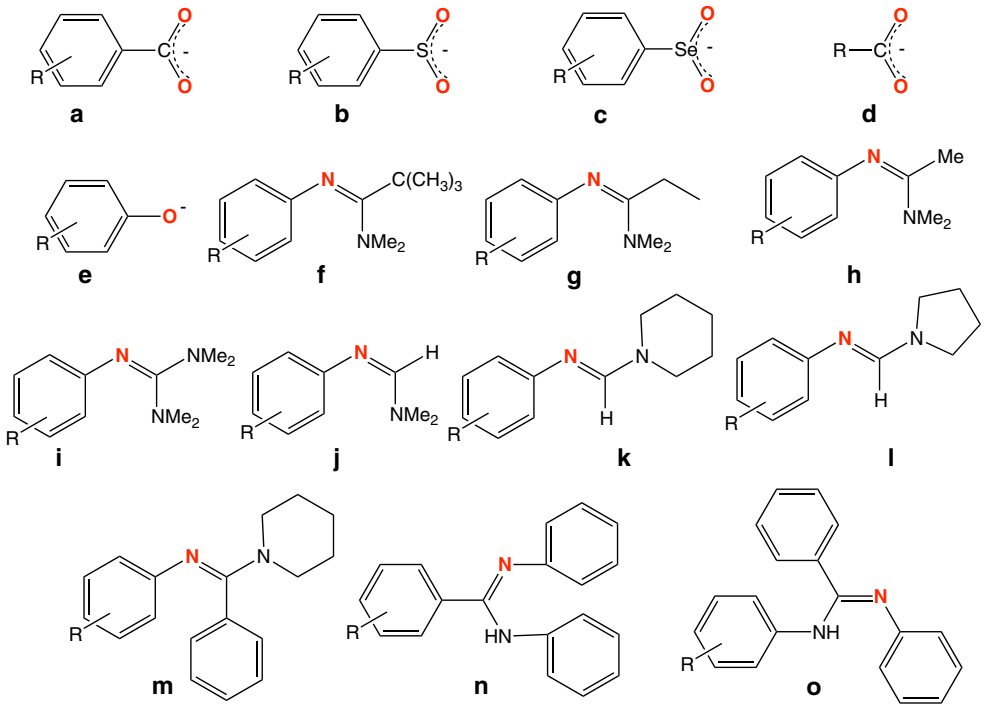

Fig. 2 15 groups (A-O) of molecules investigated in this work, different substituents R are defined in Table 1. The key atoms that are involved in the dissociation reaction and that are used for the definition of the representative local vibrational frequency pairs p(ω_1,ω_2) are shown in *red*

different compounds belonging to one of the groups A-O, summarized in Table 1. We evaluate the accuracy of the calculated pK_a values via the correlation Eq. 1 and applying cross-validation and *t* test (see below) we analyze the predictive nature of these correlations. In the last section, we summarize our results and draw conclusions.

Computational details

Local vibrational modes

The basic equation of vibrational spectroscopy [69] is defined as

$$\mathbf{f}^{\mathbf{x}}\mathbf{L} = \mathbf{M}\mathbf{L}\Lambda \qquad (2)$$

where $\mathbf{f}^{\mathbf{x}}$ is the force constant matrix expressed in Cartesian coordinates. $\mathbf{M}$ is the mass matrix. The diagonal eigenvalue matrix Λ contains N_{vib} vibrational eigenvalues $4\pi^2c^2\nu_\mu^2$ (with $\mu = 1, ..., N_{vib}$ and $N_{vib} = 3N - TR$; TR = 6 for nonlinear molecules and 5 for linear molecules) and TR zero eigenvalues corresponding to translations and rotations of the molecule. The harmonic vibrational frequencies ν_μ are given in cm^{-1} and c is the speed of light. The (3N x 3N) dimensional $\mathbf{L}$ matrix collects the N_{vib} normal vibrational mode vectors $\mathbf{l}_\mu$ and TR mode vectors corresponding to translations and rotations.

Expressing Eq. 2 in internal coordinates $\mathbf{q}$ leads to the Wilson GF formalism with [69, 70]:

$$\mathbf{F}^{\mathbf{q}}\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}\Lambda \qquad (3)$$

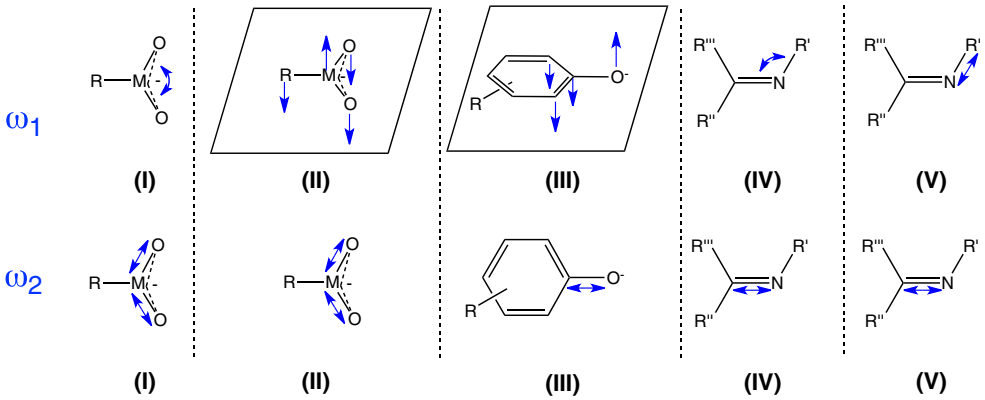

Fig. 3 Five different categories I, II, III, IV, and V of representative local vibrational frequency pairs p(ω_1,ω_2) used in this work

Table 1 Substituent R, local vibrational frequencies ω_1 and ω_2 used in Eq. 1, experimental and predicted pK_a values for all molecules of groups A-O, (180 in total). Diff is the difference between the experimental and predicted pK_a value

Group	Substituent R	ω_1 (cm^{-1})	ω_2 (cm^{-1})	pK_a (Experimental)	pK_a (Predicted)	Diff
Group A (category I)						
A-1	H	1560	899	4.20	4.23	0.03
A-2	m-Br	1568	895	3.81	3.80	−0.01
A-3	m-CN	1571	896	3.60	3.59	−0.01
A-4	m-Cl	1567	896	3.84	3.81	−0.03
A-5	m-F	1566	897	3.86	3.90	0.04
A-6	m-Me	1560	899	4.25	4.27	0.02
A-7	m-NO_2	1572	895	3.46	3.42	−0.04
A-8	m-OH	1563	898	4.08	4.09	0.01
A-9	m-OMe	1563	898	4.10	4.07	−0.03
A-10	p-Br	1565	898	3.96	3.94	−0.02
A-11	p-CN	1570	896	3.55	3.58	0.03
A-12	p-Cl	1565	898	4.00	3.99	−0.01
A-13	p-F	1562	899	4.15	4.15	0.00
A-14	p-Me	1559	899	4.37	4.39	0.02
A-15	p-NMe_2	1555	899	4.98	4.99	0.01
A-16	p-NO_2	1572	895	3.43	3.49	0.06
A-17	p-OH	1558	899	4.47	4.45	−0.02
A-18	p-OMe	1558	899	4.50	4.46	−0.04
Group B (category I)						
B-1	H	1078	534	2.97	2.96	−0.02
B-2	m-Cl	1087	528	2.26	2.26	0.01
B-3	m-NO_2	1084	516	1.67	1.67	0.00
B-4	p-Br	1083	530	2.51	2.44	−0.07
B-5	p-Cl	1082	532	2.51	2.58	0.07
B-6	p-Me	1078	536	3.12	3.13	0.01
B-7	p-NO_2	1089	501	1.43	1.43	0.00
B-8	p-OMe	1078	543	3.84	3.83	−0.01
Group C (category II)						
C-1	H	819	367	4.79	4.72	−0.07
C-2	m-Br	823	366	4.43	4.45	0.02
C-3	m-Cl	823	366	4.47	4.46	−0.01
C-4	m-F	821	360	4.34	4.30	−0.04
C-5	m-Me	818	367	4.80	4.81	0.01
C-6	m-NO_2	827	357	4.07	4.03	−0.04
C-7	m-OMe	819	361	4.65	4.68	0.03
C-8	p-Br	821	366	4.50	4.48	−0.02
C-9	p-Cl	820	367	4.48	4.51	0.03
C-10	p-F	819	369	4.50	4.47	−0.03
C-11	p-Me	818	368	4.88	4.85	−0.03
C-12	p-NO_2	824	359	4.00	4.07	0.07
C-13	p-OMe	817	369	4.65	4.73	0.08
Group D (category II)						
D-1	$ArCH_2CH_2$	1571	616	4.66	4.66	0.00
D-2	CH_3	1568	512	4.76	4.74	−0.02
D-3	C_2H_5	1569	550	4.87	4.84	−0.03
D-4	$HC{\equiv}CCH_2$	1591	530	3.32	3.32	0.00
D-5	iPr	1566	595	4.86	4.94	0.08

Table 1 (continued)

Group	Substituent R	ω_1 (cm^{-1})	ω_2 (cm^{-1})	pK_a (Experimental)	pK_a (Predicted)	Diff
D-6	n-C_4H_9	1570	565	4.84	4.84	0.00
D-7	n-C_5H_{11}	1570	436	4.88	4.87	−0.01
D-8	n-C_6H_{13}	1570	440	4.89	4.91	0.02
D-9	n-C_9H_{19}	1570	549	4.89	4.85	−0.04
D-10	n-Pr	1569	546	4.76	4.84	0.08
D-11	tBu	1565	602	5.05	4.99	−0.06
Group E (category III)						
E-1	H	1535	731	9.98	10.02	0.04
E-2	m-Br	1556	724	9.03	8.99	−0.04
E-3	m-CH_3	1536	730	10.08	9.97	−0.11
E-4	m-CN	1558	721	8.61	8.74	0.13
E-5	m-Cl	1553	727	9.02	9.16	0.14
E-6	m-F	1548	734	9.28	9.31	0.03
E-7	m-NH_2	1537	735	9.87	9.84	−0.03
E-8	m-NO_2	1559	716	8.40	8.32	−0.08
E-9	m-OCH_3	1542	733	9.65	9.64	−0.01
E-10	m-OH	1544	732	9.44	9.57	0.13
E-11	p-Br	1550	726	9.36	9.27	−0.09
E-12	p-CH_3	1531	727	10.14	10.11	−0.03
E-13	p-CN	1583	729	7.95	7.81	−0.14
E-14	p-Cl	1546	726	9.38	9.45	0.07
E-15	p-F	1526	724	9.95	10.15	0.20
E-16	p-NH_2	1531	726	10.30	10.05	−0.25
E-17	p-NO_2	1599	726	7.15	7.21	0.06
E-18	p-OCH_3	1519	720	10.21	10.07	−0.14
E-19	p-OH	1514	718	9.96	10.07	0.11
Group F (category V)						
F-1	H	1572	1188	8.03	8.09	0.06
F-2	m-Br	1562	1193	6.81	6.77	−0.04
F-3	m-CH_3	1574	1187	8.24	8.21	−0.03
F-4	m-Cl	1563	1192	7.01	6.97	−0.04
F-5	m-NO_2	1553	1196	5.86	5.88	0.02
F-6	m-OCH_3	1571	1187	7.91	7.86	−0.05
F-7	m-OC_2H_5	1573	1187	7.93	8.11	0.18
F-8	p-Br	1566	1192	7.20	7.27	0.07
F-9	p-CH_3	1573	1188	8.43	8.28	−0.15
F-10	p-Cl	1567	1193	7.19	7.24	0.05
F-11	p-I	1566	1193	7.17	7.11	−0.06
F-12	p-NO_2	1550	1199	5.18	5.18	0.00
F-13	p-OCH_3	1575	1187	8.62	8.56	−0.06
F-14	p-OC_2H_5	1576	1186	8.63	8.67	0.04
Group G (category V)						
G-1	H	1579	1192	8.30	8.19	−0.11
G-2	m-Br	1571	1197	7.13	7.13	0.00
G-3	m-CH_3	1581	1190	8.38	8.44	0.06
G-4	m-Cl	1572	1197	7.15	7.21	0.06
G-5	m-NO_2	1559	1202	6.06	6.05	−0.01
G-6	m-OCH_3	1579	1192	8.13	8.17	0.04

Table 1 (continued)

Group	Substituent R	ω_1 (cm^{-1})	ω_2 (cm^{-1})	pK_a (Experimental)	pK_a (Predicted)	Diff
G-7	m-OC_2H_5	1580	1192	8.16	8.29	0.13
G-8	p-Br	1575	1198	7.43	7.46	0.03
G-9	p-CH_3	1580	1190	8.44	8.36	−0.08
G-10	p-Cl	1575	1197	7.55	7.48	−0.07
G-11	p-I	1574	1198	7.36	7.33	−0.03
G-12	p-NO_2	1564	1205	5.65	5.66	0.01
G-13	p-OCH_3	1579	1186	8.94	9.00	0.06
G-14	p-OC_2H_5	1580	1187	8.89	8.80	−0.09
Group H (category V)						
H-1	H	1578	1187	8.32	8.36	0.04
H-2	m-Br	1568	1194	7.19	7.19	0.00
H-3	m-CH_3	1579	1185	8.41	8.56	0.15
H-4	m-Cl	1569	1193	7.25	7.26	0.01
H-5	m-OCH_3	1577	1188	8.22	8.13	−0.09
H-6	m-OC_2H_5	1578	1188	8.26	8.31	0.05
H-7	p-Br	1572	1192	7.55	7.54	−0.01
H-8	p-CH_3	1580	1186	8.65	8.53	−0.12
H-9	p-Cl	1573	1192	7.65	7.66	0.01
H-10	p-NO_2	1558	1203	5.69	5.69	0.00
H-11	p-OCH_3	1580	1183	8.96	8.91	−0.05
H-12	p-OC_2H_5	1581	1183	8.90	8.91	0.01
Group I (category IV)						
I-1	H	1535	567	11.52	11.55	0.03
I-2	m-CH_3	1535	564	11.74	11.71	−0.03
I-3	m-Cl	1520	564	10.55	10.52	−0.03
I-4	m-OCH_3	1531	564	11.44	11.47	0.03
I-5	m-OC_2H_5	1537	558	11.38	11.38	0.00
I-6	p-CH_3	1538	565	11.94	11.85	−0.09
I-7	p-Cl	1526	564	10.98	11.04	0.06
I-8	p-OCH_3	1544	570	12.16	12.16	0.00
I-9	p-OC_2H_5	1543	565	12.08	12.13	0.05
Group J (category IV)						
J-1	H	1620	524	7.45	7.31	−0.14
J-2	m-Br	1614	518	6.45	6.42	−0.03
J-3	m-CH_3	1621	528	7.63	7.65	0.02
J-4	m-Cl	1615	521	6.50	6.62	0.12
J-5	m-OCH_3	1621	523	7.45	7.51	0.06
J-6	m-OC_2H_5	1605	600	7.45	7.47	0.02
J-7	p-Br	1614	523	6.69	6.61	−0.08
J-8	p-CH_3	1608	598	7.75	7.73	−0.02
J-9	p-Cl	1615	526	6.84	6.88	0.04
J-10	p-NO_2	1584	590	5.25	5.25	0.00
J-11	p-OCH_3	1621	539	7.91	7.95	0.04
J-12	p-OC_2H_5	1620	538	7.83	7.78	−0.05
Group K (category IV)						
K-1	H	1601	595	8.69	8.60	−0.09
K-2	m-Br	1590	600	7.70	7.70	0.00
K-3	m-CH_3	1602	596	8.74	8.83	0.09

Table 1 (continued)

Group	Substituent R	ω_1 (cm^{-1})	ω_2 (cm^{-1})	pK_a (Experimental)	pK_a (Predicted)	Diff
K-4	m-Cl	1592	594	7.83	7.84	0.01
K-5	m-OCH_3	1599	598	8.47	8.52	0.05
K-6	p-Br	1593	597	7.95	7.95	0.00
K-7	p-CH_3	1603	601	9.10	9.10	0.00
K-8	p-Cl	1595	599	8.09	8.06	−0.03
K-9	p-OCH_3	1604	597	9.38	9.34	−0.04
K-10	p-OC_2H_5	1606	592	9.10	9.11	0.01
Group L (category V)						
L-1	H	1600	1194	9.31	9.41	0.10
L-2	m-Br	1593	1200	8.46	8.40	−0.06
L-3	m-CH_3	1601	1193	9.53	9.56	0.03
L-4	m-Cl	1593	1201	8.52	8.59	0.07
L-5	m-OCH_3	1599	1192	9.22	9.22	0.00
L-6	p-Br	1594	1201	8.75	8.76	0.01
L-7	p-CH_3	1602	1193	9.75	9.61	−0.14
L-8	p-Cl	1595	1200	8.85	8.82	−0.03
L-9	p-OCH_3	1605	1191	9.97	9.98	0.01
L-10	p-OC_2H_5	1605	1192	9.70	9.71	0.01
Group M (category V)						
M-1	H	1564	1189	7.45	7.51	0.06
M-2	m-Br	1552	1195	6.38	6.38	0.00
M-3	m-CH_3	1564	1187	7.72	7.66	−0.06
M-4	m-Cl	1554	1195	6.45	6.46	0.01
M-5	m-OCH_3	1563	1189	7.35	7.43	0.08
M-6	m-OC_2H_5	1564	1189	7.58	7.56	−0.02
M-7	p-Br	1557	1195	6.65	6.67	0.02
M-8	p-CH_3	1565	1187	7.90	7.83	−0.07
M-9	p-Cl	1558	1194	6.77	6.73	−0.04
M-10	p-OCH_3	1567	1185	8.19	8.21	0.02
M-11	p-OC_2H_5	1567	1185	8.17	8.19	0.02
Group N (category V)						
N-1	H	1593	1187	14.65	14.56	−0.09
N-2	m-Br	1596	1186	13.81	13.85	0.04
N-3	m-NO_2	1596	1186	13.38	13.37	−0.01
N-4	p-Br	1594	1187	14.05	14.10	0.05
N-5	p-CH_3	1591	1187	14.92	14.96	0.04
N-6	p-NO_2	1596	1185	13.19	13.17	−0.02
N-7	p-OCH_3	1588	1188	15.27	15.27	0.00
Group O (category V)						
O-1	H	1594	1173	6.15	6.16	0.01
O-2	m-Br	1591	1178	5.25	5.26	0.01
O-3	m-CH_3	1596	1171	6.36	6.51	0.15
O-4	m-Cl	1591	1178	5.23	5.28	0.05
O-5	m-NO_2	1589	1184	4.24	4.29	0.05
O-6	m-OCH_3	1598	1173	6.04	6.01	−0.03
O-7	p-Br	1590	1178	5.42	5.38	−0.04
O-8	p-CH_3	1595	1172	6.52	6.39	−0.13

Table 1 (continued)

Group	Substituent R	ω_1 (cm^{-1})	ω_2 (cm^{-1})	pK_a (Experimental)	pK_a (Predicted)	Diff
O-9	p-Cl	1590	1178	5.46	5.46	0.00
O-10	p-F	1591	1174	5.90	6.08	0.18
O-11	p-I	1590	1179	5.43	5.26	−0.17
O-12	p-OCH_3	1593	1170	6.78	6.70	−0.08

where

$$\mathbf{G} = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^{\dagger} \tag{4}$$

$\mathbf{F^q}$ is the force constant matrix expressed in terms of internal coordinates **q**. Each normal mode vector $\mathbf{d}_\mu$ represents a column vector of the (N_{vib}xN_{vib}) dimensional **D** matrix, which is defined as the product **BL**. The rectangular (N_{vib}x3N) dimensional **B** matrix contains the first derivatives of the internal coordinates with regard to the Cartesian coordinates, thus connecting both coordinate systems, and the (N_{vib}xN_{vib}) dimensional matrix **G** is the Wilson **G** matrix. [69, 70].

The normal mode frequencies ν_μ are coupled caused by electronic coupling (off-diagonal elements of the $\mathbf{F^q}$ matrix, which can be eliminated via the Wilson GF formalism) and mass-coupling (caused by the off-diagonal elements of the Wilson **G** matrix) [69, 70]. However, for our correlation, we need local vibrational modes focusing on the influence of the substituent R on the electronic structure of the conjugated base. Konkoli and Cremer [68] solved this problem by introducing the concept of local vibrational modes. They solved the mass-decoupled Euler–Lagrange equation by setting all the atomic masses to zero except those of the molecular fragment (e.g., bond, angle, or dihedral, etc.) carrying out a localized vibration. As shown in their original work, the change in the local displacement of a specific internal coordinate is equivalent to an adiabatic relaxation of the molecule [68]. For any molecular fragment associated with an internal coordinate q_n, the corresponding local mode vector $\mathbf{a}_n$ is given by

$$\mathbf{a}_n = \frac{\mathbf{K}^{-1}\mathbf{d}_n^{\dagger}}{\mathbf{d}_n\mathbf{K}^{-1}\mathbf{d}_n^{\dagger}} \tag{5}$$

where $\mathbf{d}_n$ contrary to $\mathbf{d}_\mu$ is a row vector of matrix **D**. Matrix **K** is the diagonal matrix of force constants k_Q, expressed in normal coordinates Q_μ with

$$\mathbf{F^Q} = \mathbf{K} = \mathbf{L}^{\dagger}\mathbf{f}^{x}\mathbf{L} \tag{6}$$

resulting from the Wilson GF formalism shown in Eqs. 3 and 4.

The local mode force constant k_n^a corresponding to local mode $\mathbf{a}_n$ is obtained by

$$k_n^a = \mathbf{a}_n^{\dagger}\mathbf{K}\mathbf{a}_n \tag{7}$$

The local vibrational frequency ω_n^a corresponding to local mode $\mathbf{a}_n$ is obtained by

$$(\omega_n^a)^2 = \frac{1}{4\pi^2 c^2} k_n^a G_{nn} \tag{8}$$

in which G_{nn} is a diagonal element of the Wilson **G**-matrix and corresponds to the reduced mass of the local mode $\mathbf{a}_n$ [68].

Local vibrational modes have successfully been applied to quantify weak chemical interactions such as hydrogen bonding [71–76], halogen bonding [77–79], pnicogen bonding [80–82], chalcogen bonding [83] and tetrel bonding [84], and to derive new chemical descriptors such as a new aromaticity index [85–87] or a generalized Tolman electronic parameter [88–90], as well as for the derivation of a generalized Badger Rule [91] and several others new concepts [91–97].

Geometry optimizations and normal mode calculations for all molecules investigated in this work were performed in the gas phase with the ωB97XD density functional [98] using the Gaussian 16 quantum chemistry program [99, 100]. For members of groups B, D, E, F, G, H, I, J, K, L ,M, N, and O (see Fig. 2) Dunning's cc-pVTZ basis set [101–103] was applied, while for members of group A and C (see Fig. 2) Pople's 6-31++G(d,p) basis set [104–106] turned out to be the best choice. Experimental pK_a values were taken from ref [1, 62]. There are no reliable experimental pK_a values for sulphinic acid available. Therefore, we used computed pK_a values instead [107]. The local mode analysis was performed with the COLOGNE2017 program package [108].

Results and discussion

Correlation of pK_a and local vibrational frequencies

Vibrational frequencies are second-order response properties [109] and therefore they are sensitive to any changes in the electronic structure of a molecule. First it was important to determine the local vibrational modes that best reflect the influence of the substituent R on the dissociation reaction for each molecule investigated in this work.

Depending on the molecule, there exist two or three vibrational modes near the dissociating proton that will dominate the pK_a value. As there are some redundancies between stretching, bending, and pyramidalization modes at a specific molecular site with regard to the electronic structure change during the vibration, one has to remove either one in order to obtain a robust model. Testing ${}^3C_2 = 3$ different combinations, we identified the set of parameters that performs best and leads to the desired model. This led to the five different categories of local vibrational mode pairs $p(\omega_1,\omega_2)$ shown in Fig. 3 used in this study. For category I molecules, a pair of local bending and stretching vibrations turned out to be most sensitive; for category II and category III molecules, a pair of local pyramidalization and stretching vibrations; for category IV molecules, a pair of local bending and stretching vibrations, and for category V molecules, two local stretching vibrations. The two local stretching vibrations for category I and II molecules were averaged. Only one of the substituent (R or R' or R" or R"') was changed at a time. The 15 groups A-O of molecules used for the correlation comprise the conjugate bases of (A) benzoic acid; (B) benzene sulphinic acid; (C) benzeneselenic acids; (D) alkyl carboxylic acids; (E) phenols; (F) protonated N^1, N^1-dimethyltertiarybutylamidine; (G) protonated N^1, N^1-dimethylethylamidine; (H) protonated N^1, N^1-dimethylethylamine; (I) protonated N^2,-alkyltetramethylguanidine; (J) protonated N^1, N^1-dimethylformamidine; (K) protonated N^1, N^1-(pentamethylene-1,5)-formamidine; (L) protonated N^1, N^1-(butamethylene-1,4)-formamidine; (M) protonated N^1, N^1-(pentamethylene-1,5)-benzamidine; (N) protonated *N*,*N*'-diphenyl-benzamidine; (O) protonated N,N'-diphenyl-benzamidine, shown in Fig. 2.

In the following, the results of our investigation are discussed for each individual group, referring to the data in Table 1, which includes for each individual compound the substituent R, the local vibrational mode frequencies ω_1 and ω_2 used in Eq. 1, experimental pK_a and predicted pK_a values, as well as the difference $DIFF$ between experimental and predicted pK_a values. The correlation protocol for groups A-O is presented in Table 2 including for each group the number of molecules, the root mean square error (RMSE), the mean average error (MAE), the mean average derivation (MAD), and the R^2 value, as a statistical measure of how close the data are to the fitted regression line.

Group A consists of 18 different compounds, which all fall into the category I of local mode frequency pairs $p(\omega_1,\omega_2)$, as shown in Table 1. The substituents vary from electron donating groups (EDG) to electron withdrawing groups (EWG), which accordingly change the electronic structure of the parent compound (R = H) over a wide range, as such influencing the pK_a value by stabilizing/destabilizing the corresponding conjugate base. We obtained a strong correlation with R^2 greater than 0.99 and a MAE of 0.025 pK_a units for this group, reflecting the fact that the pK_a values of group A molecules are dominated by substituent effects. It has to be noted that the error of the experimental data is as high as ± 0.09 pK_a units as reported in ref [1], which may amplify the predicted error due to quadratic fitting (QFE). **Group B** consists of eight different compounds and shows a MAD of 0.578. Despite this large MAD value, we observed a strong correlation with R^2 of 1.0 and a MAE of 0.02. A similarly strong correlation was found for **group C** and **group D** molecules, again reflecting that the pK_a values of these groups members are dominated by substituent effects. **Group E** shows the two largest deviations between calculated and experimental pK_a values of all molecules investigated in this work, namely for R = p-F and R = p-NH_2. Apart from the QFE, another reason for these deviations might be the model chemistry used. The optimal level of theory and basis set for each molecule varies according to the type of substituent R involved. We tested different levels of theory and basis sets, which led to somewhat different outliers, (see supporting information for a detailed report).

Group F, G and **H** members are characterized by a substituent R" at the C atom of the C=N bond with a decreasing number of C atoms, e.g., **group F**: -$C(CH_3)_3$, **group G**: -C_2H_5 and **group H** -CH_3, respectively. The number of carbon atoms in R" determine changes in the electronic structure caused by inductive effects, which is reflected by the local C=N double bond stretching frequency. Largest inductive effects were found for **group G** and **group F** molecules, in particular for molecules G-7 and F-7, substituent m-OC_2H_5, the outliers of these groups. In **group F** also the -CH_3 functional group F-9 led to an outlier, caused by an increased resonance effect, which we also found for molecule H-8 of **Group H**. **Group I** molecules show a similar deviations as found for **group H** members, due to resonance effects. **Group J** and **group K** show the largest outliers for the parent compound (R = H), molecules J-1 and K-1, respectively. As all other substituents in these two groups are either strong EDG or EWG functional groups, R = H deviates from the general group pattern. In **Group L** and **group M** the p-CH_3 functional group leads again to the largest outliers. Due to hyperconjugation, the p-CH_3 substituent increases the electron density at the *para* position, and in this way strongly influences the electronic structure. **Group O** has several outliers. Nevertheless, we observe a strong correlation with R^2 of almost 0.98. The outliers in this group can be attributed to both the QFE and the model chemistry used.

The electronic structure is affected by resonance, inductive effects or by hyperconjugation. It is remarkable to see that even changes in substituents at distance from

Table 2 Correlation protocol for each group A-O

Group	Number of molecules	RMSE[a]	MAE[b]	MAD[c]	R^2 [d]
Group A	18	0.028	0.025	0.311	0.995
Group B	8	0.035	0.022	0.578	0.998
Group C	13	0.044	0.038	0.192	0.970
Group D	11	0.041	0.031	0.261	0.991
Group E	19	0.114	0.096	0.637	0.981
Group F	14	0.077	0.060	0.812	0.994
Group G	12	0.067	0.056	0.779	0.995
Group H	12	0.066	0.045	0.712	0.994
Group I	9	0.043	0.035	0.398	0.992
Group J	12	0.065	0.052	0.628	0.992
Group K	10	0.047	0.033	0.497	0.993
Group L	10	0.062	0.046	0.449	0.985
Group M	11	0.043	0.036	0.557	0.995
Group N	7	0.044	0.035	0.656	0.996
Group O	12	0.098	0.075	0.560	0.979

[b] $MAE = \frac{1}{n}\sum|y_j^e - y_j^p|$, [a] $RMSE = \sqrt{\frac{1}{n}\sum(y_j^e - y_j^p)^2}$, [c] $MAD = \frac{1}{n}\sum|y_j^e - \bar{y_j}^e|$ y equals the pK_a; superscript e denotes experimental, p predicted; $\bar{y_j}^e$ is the mean of the experimental pK_a values, j equals the number of molecules in a group.

[d] R^2 is the statistical measure for how close the data points are to the fitted regression line

the protonation center can be captured by the two local vibrational reference frequencies ω_1, and ω_2 involving the protonation center. It is also noteworthy that although all predicted pK_a values were derived from gas phase calculations, while the experimental pK_a values were measured in solution, the local vibrational modes calculated in the gas phase capture already the important electronic structure changes influencing the pK_a values. This allows for a quick check of the pK_a values for new group members with different substituents R.

General trends

As reflected by the data in Table 2, each group shows a very strong correlation with an R^2 value greater than 0.97 and a MAE of less than 0.09. By inspection of the correlation constants in Eq. 1 (see supporting information for further details), one can easily see that there is however no relation between the different groups, and hence an overall correlation is not meaningful. For each particular group, the changes of the electronic structure upon substitution leading to a change in the molecular acidity are dominated by the substituent. Since the local vibrational modes can capture the sensitive changes in the electronic structure of conjugate base, we obtain a significant correlation between pK_a values and local vibrations. Figure 4 shows the predicted and experimental pK_a values. The results for MAE, RMSE, and R^2, which are 0.05, 0.07, and 0.999, respectively, confirm our hypothesis that there exists a strong correlation between pK_a and local vibrational frequencies. In order to check if this correlation results predominantly from an electronic effect, we also correlated local mode force constants (Eq. 7) with experimental pK_a values, because the local mode force constants are free from any mass effects. We observed similar results as for the local vibrational frequencies, (detailed information can be found in the supporting information).

Validation of the correlation

In order to test the predictive power of our new correlation model, we performed a k-fold cross-validation for molecules similar to group E members, including 33 compounds of singly and doubly substituted benzenes, (shown in the supporting information). Cross-validation is a reliable means to check the efficiency of predicting models. In a k-fold cross-validation, data is randomly distributed to k splits as shown in Fig. 5. A model is then trained for k-1 splits and tested for the left out split. Any extra data point(s) is randomly assigned to one of the splits. This is repeated k times and the final error (MAE or RMSE) is averaged to get k-fold cross-validation error. We performed a four-fold cross-validation for this group and obtained a cross-validation MAE of 0.05, which equals the actual MAE for this group. This proves that our model is predictive in nature and can be used to predict pK_a values for any new group member with a different substituent R for any group of molecules, provided we have access to sufficient

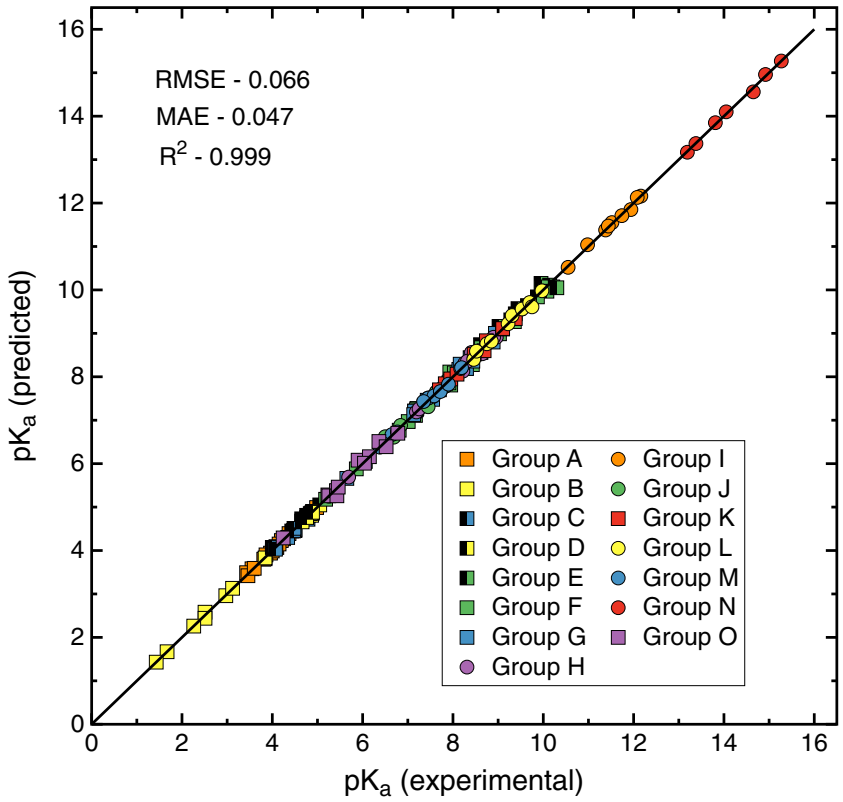

Fig. 4 Correlation between experimental and predicted pK_a values

experimental pK_a values and the representative local mode pairs $p(\omega_1,\omega_2)$ for a variety of substituted molecules of this group, so that Eq. 1 can be solved.

The *t* test is a measure of statistical significance between two different data distributions, which can be computed as

$$t_{value} = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{9}$$

where, $\bar{x}$, s and n are mean, standard deviation and variance of the data. Subscript 1 and 2 denotes the first data set and the second data set, respectively.

We calculated t_{value} for all pairs of coefficients (Group X coefficients (data set 1) with Group Y coefficients (data set 2), where $X, Y \in [A, B...O]$). The degrees of freedom can be calculated as

$$df = l_1 + l_2 - 2 \tag{10}$$

where, l_1, l_2 are the number of data points in data set 1 and data set 2, respectively. With the degrees of freedom of 10 (6+6-2) and 95% confidence, we keep the null hypothesis (the coefficients are statistically similar) if the t_{value} is less than 2.228 (two tailed t-distribution table). The maximum t_{value} was 1.5465, which confirms the null hypothesis, i.e.,

Fig. 5 k-fold cross-validation. E is the error in each iterations

there is no statistically significant difference between each of the two models.

Conclusions and outlook

Our investigation shows that the changes in electronic structure caused by different substituent R are captured by local vibrational frequencies related to the conjugate base. We verified that there exists a relationship between the local vibrational modes and the pK_a of a molecule provided substituent effects play the major role in determining the pK_a value. Fifteen different groups of molecules with various substituents (a total of 180 molecules) were investigated and showed strong correlations. The slopes and intercepts obtained in the correlation model were different for each group, which reflects that pK_a values are system dependent. However, within a group, the system dependency of pK_a value is removed and we can predict the pK_a value of any new group members with different substituents R.

Using cross-validation and *t* test, we showed that our model is predictive in nature and offers a lot of potential for the pK_a prediction for other groups of molecules, which is currently under investigation, including the following two examples. (i) Huge efforts are devoted to capture CO and CO_2 and to store these gases in an efficient manner [110–112]. Amine solutions play an important role in this regard [113–115]. The pK_a values of these amines are important indicators for their adsorption efficiency [116, 117]. A wide range of pK_a values have been collected for amine species [61, 117–120], therefore we can apply our correlation model to estimate the pK_a values of novel amine substituted amines with a potentially higher adsorption efficiency. (ii) The ionic form of a weak acid/base varies across a range of pH values. This is important in physiological systems, in which the ionization state affects the rate of diffusion across membranes. The pK_a influences permeability, protein binding, solubility, etc., which in turn affects absorption, metabolism, excretion, etc., of potential drug candidates [121–126]. Due to this connection, the pK_a plays an important role in engineering optimal pharmacokinetic characteristics. A large number of drug candidates have already been studied [121, 122] leading to a wealth of experimental pK_a values, serving as the basis for our correlation model.

In summary, our new correlation model constitutes a powerful link between the well-known Hammett equation and vibrational spectroscopy. From a practical perspective, one needs only to calculate the local vibrational modes in the gas phase. The correlation with experimental pK_a values removes the system dependency and leads to an equation that allows a quick check of the pK_a value of a new compound with a different substituent R, avoiding any extensive pK_a calculation. This makes our approach the perfect tool for the engineering of compounds with a specific molecular acidity regulated by substituent effects.

Acknowledgements This work was financially supported by National Science Foundation Grants CHE 1464906. We thank SMU for generous supercomputer resources.

References

1. Cao X, Rong C, Zhong A, Tian LiuS (2018) Molecular acidity: an accurate description with information-theoretic approach in density functional reactivity theory. J Comp Chem 39:117–129
2. Cramer CJ, Truhlar DG (1999) Implicit solvation models: Equilibria, structure, spectra, and dynamics. Chem Rev 99(8):2161–2200
3. Alongi KS, Shields GC (2010) Theoretical calculations of acid dissociation constants: A review article. Annu Rep Comput Chem 6:113–138
4. Sastre S, Casasnovas R, Munozac F, Frau J (2016) Isodesmic reaction for accurate theoretical K_a calculations of amino acids and peptides. Phys Chem Chem Phys 18:11202–11212
5. Casasnovas R, Ortega-Castro J, Frau J, Donoso J, Munoz F (2014) Theoretical pK_a calculations with continuum model solvents, alternative protocols to thermodynamic cycles. Int J Quantum Chem 114(20):1350–1363
6. Lanevskij K, Japertas P, Didziapetris R, Petrauskas A (2009) Ionization-specific prediction of blood–brain permeability. J Pharm Sci 98(1):122–134
7. Dreanic MP, Edge CM, Tuttle T (2017) New insights into the catalytic mechanism of aldose reductase: a QM/MM study. ACS Omega 2:5737–5747
8. Rajapakse HA, Nantermet PG, Selnick HG, Barrow JC, McGaughey GB, Munshi S, Lindsley SR, Young MB, Ngo PL, Holloway MK, Lai MT, Espeseth AS, Shi XP, Colussi D, Pietrak B, Crouthamel MC, Tugusheva K, Huang Q, Xu M, Simon AJ, Kuo L, Hazuda DJ, Graham S, Vacca JP (2010) Sar of tertiary carbinamine derived bace1 inhibitors: role of aspartate ligand amine pK_a in enzyme inhibition. Bioorganic Med Chem Lett 20(6):1885–1889
9. Zhang J, Liu Z, Lian P, Qian J, Li X, Wang L, Fu W, Chen L, Wei X, Li C (2016) Selective imaging and cancer cell death via pH-switchable near-infrared fluorescence and photothermal effects. Chem Sci 7(9):5995–6005
10. Aiken GR, Hsu-Kim H, Ryan JN (2011) Influence of dissolved organic matter on the environmental fate of metals, nanoparticles, and colloids. Environ Sci Technol 45(8):3196–3201
11. Weng L, Temminghoff EJM, Lofts S, Tipping E, Van Riemsdijk WH (2002) Complexation with dissolved organic matter and solubility control of heavy metals in a sandy soil. Environ Sci Technol 36(22):4804–4810
12. Rao B, Simpson C, Lin H, Liang L, Gu B (2014) Determination of thiol functional groups on bacteria and natural organic matter in environmental systems. Talanta 119:240–247
13. Cheng J, Sprik M (2010) Acidity of the aqueous rutile TiO_2(110) surface from density functional theory-based molecular dynamics. J Chem Theory Comput 6(3):880–889

14. Bryantsev VS (2013) Predicting the stability of aprotic solvents in Li-Air batteries: pK_a calculations of aliphatic C-H acids in dimethyl sulfoxide. Chem Phys Lett 558:42–47. supplement C
15. Gallus DR, Wagner R, Wiemers-Meyer S, Winter M, Cekic-Laskovic I (2015) New in-sights into the structure-property relationship of high-voltage electrolyte components for lithium-ion batteries using the pK_a value. Electrochimica Acta 184:410–416. supplement C
16. Lau YH, Clegg JK, Price JR, Macquart RB, Todd MH, Rutledge PJ (2018) Molecular switches for any pH: a systematic study of the versatile coordination behaviour of cyclam scorpionands. Chem Eur J 24:1573–1585
17. Yu D, Du R, Xiao JC, Xu S, Rong C, Liu S (2018) Theoretical study of pK_a values for trivalent rare-earth metal cations in aqueous solution. J Phys Chem A 122:700–707
18. Selwa E, Kenney IM, Beckstein O, Iorga BI (2018) SAMPL6: calculation of macroscopic pK_a values from ab initio quantum mechanical free energies. J Computer-Aided Mol Design 32:1203–1216
19. 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 Computer-Aided Mol Design 32:1139–1149
20. Lian P, Johnston RC, Parks JM, Smith JC (2018a) Quantum chemical calculation of pK_as of environmentally relevant functional groups: carboxylic acids, amines, and thiols in aqueous solution. J Phys Chem A 122:4366–4374
21. Hu J, Zhao S, Geng W (2018) Accurate pK_a computation using matched interface and boundary (MIB) method-based Poisson–Boltzmann solver. Commun Comput Phys 23:520–539
22. Romero R, Salgado PR, Soto C, Contreras D, Melin V (2018) An experimental validated computational method for pK_a determination of substituted 1,2-dihydroxybenzenes. Front Chem 6:208
23. Li M, Zhang H, Chen B, Wu Y, Guan L (2018) Prediction of pK_a values for neutral and basic drugs based on hybrid artificial intelligence methods. Sci Rep 8:3991
24. Shields G, Seybold P (2014) Computational approaches for the prediction of pK_a values. CRC Press, Boca Raton
25. Reijenga J, van Hoof A, van Loon A, Teunissen B (2013) Development of methods for the determination of pK_a values. Anal Chem Insights 8:53–71
26. Jorgensen WL, Briggs JM, Gao JA (1987) A priori calculations of pK_a's for organic compounds in water. The pK_a of ethane. J Am Chem Soc 109:6857
27. Potter MJ, Gilson MK, McCammon JA (1994) Small molecule pK_a prediction with continuum electrostatics calculations. J Am Chem Soc 116:10298
28. Matsui T, Shigeta Y, Morihashi K (2017) Assessment of methodology and chemical group dependences in the calculation of the pK_a for several chemical groups. J Chem Theory Comput 13:4791. https://doi.org/10.1021/acs.jctc.7b00587
29. Baba T, Matsui T, Kamiya K, Nakano M, Shigeta Y (2014) A density functional study on the pK_a of small polyprotic molecules. Int J Quant Chem 114:1128
30. Rossini E, Netz RR, Knapp EW (2016) Computing pK_a values in different solvents by electrostatic transformation. J Chem Theory Comput 12:3360
31. Han WG, Noodleman L (2010) Quantum cluster size and solvent polarity effects on the geometries and Mssbauer properties of the active site model for ribonucleotide reductase intermediate X: a density functional theory study. Theor Chem Acc 125(3-6):305–317
32. Haworth NL, Wang Q, Coote ML (2017) Modeling flexible molecules in solution: A pK_a case study. J Phys Chem A 121:5217
33. Sastre S, Casasnovas R, Munoz F, Frau J (2013) Isodesmic reaction for pK_a calculations of common organic molecules. Theor Chem Acc 132:1310
34. Schmidt am Busch M, Knapp EW (2004) Accurate pK_a determination for a heterogeneous group of organic molecules. ChemPhysChem 5:1513
35. Chaudry UA, Popelier PLA (2004) Estimation of pK_a using quantum topological molecular similarity descriptors: application to carboxylic acids, anilines and phenols. J Org Chem 69:233
36. Yu D, Du R, Xiao JC (2016) pK_a prediction for acidic phosphorus-containing compounds using multiple linear regression with computational descriptors. J Comput Chem 37:1668
37. Silva CO, da Silva EC, Nascimento MAC (2000) Ab initio calculations of absolute pK_a values in aqueous solution II. Aliphatic alcohols, thiols, and halogenated carboxylic acids. J Phys Chem A 104(11):2402–2409
38. Pliego JR (2003) Thermodynamic cycles and the calculation of pK_a. Chem Phys Lett 367:145–149
39. 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
40. Namazian M, Halvani S (2006) Calculations of pK_a values of carboxylic acids in aqueous solution using density functional theory. J Thermodyn Chem 38:1495–1502
41. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phy Chem B 113(18):6378–6396. https://doi.org/10.1021/jp810292n
42. Lian P, Johnston RC, Parks JM, Smith JC (2018) Quantum chemical calculation of pK_a's of environmentally relevant functional groups: Carboxylic acids, amines and thiols in aqueous solution. J Phy Chem A 122(17):4366–4374. https://doi.org/10.1021/acs.jpca.8b01751
43. Miertus S, Scrocco E, Tomasi J (1981) Electrostatic interaction of a solute with a continuum. A direct utilization of ab-initio molecular potentials for the prevision of solvent effects. Chem Phys 55(1):117–129
44. Miertus S, Tomasi J (1982) Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. Chem Phys 65(2):239–245
45. Pascual-ahuir JL, Silla E, Tunon I (1994) GEPOL: an improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. J Comput Chem 15(10):1127–1138
46. Scalmani G, Frisch MJ (2010) Continuous surface charge polarizable continuum models of solvation. I. General formalism. J Chem Phys 132(11):114110
47. Foresman JB, Keith TA, Wiberg KB, Snoonian J, Frisch MJ (1996) Solvent effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation on ab-initio reaction field calculations. J Phys Chem 100(40):16098–16104
48. Cossi M, Rega N, Scalmani G, Barone V (2003) Energies, structures, and electronic properties of molecules in solution with the c-PCM solvation model. J Comput Chem 24(6):669–681
49. Klamt A, Schuurmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc 0(5):799–805

50. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J Phys Chem 99(7):2224–2235
51. Barone V, Cossi M (1998) Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. J Phys Chem A 102(11):1995–2001
52. Car R, Parrinello M (1985) Unified approach for molecular dynamics and density functional theory. Phys Rev Lett 55:2471
53. Citra MJ (1999) Estimating the pK_a of phenols, carboxylic acids and alcohols from semi-empirical quantum chemical methods. Chemosphere 38(1):191–206
54. Czodrowski P, Dramburg I, Sotriffer CA, Klebe G (2006) Development, validation, and application of adapted PEOE charges to estimate pK_a values of functional groups in protein-ligand complexes. Proteins: Struct, Funct, Bioinf 65(2):424–437
55. Hall HK (1957) Correlation of the base strengths of amines[1]. J Am Chem Soc 79(20):5441–5444
56. Nogaj B, Dulewicz E, Brycki B, Hrynio A, Barczynski P, Dega-Szafran Z, Szafran M, Koziol P, Katritzky AR (1990) Chlorine-35 nuclear quadrupole resonance and infrared spectroscopic studies of hydrogen bonding in complexes of dichloroacetic acid with nitrogen and oxygen bases: Correlation of spectroscopic properties with proton affinity and aqueous pK_a. J Phys Chem 94(4):1279–1285
57. Charif IE, Mekelleche SM, Villemin D, Mora-Diezc N (2007) Correlation of aqueous pK_a values of carbon acids with theoretical descriptors: A DFT study. J Mol Struct: THEOCHEM 818(1-3):1–6
58. Gasque L, Medina G, Ruiz-Ramirez L, Moreno-Esparza R (1999) Cu-O stretching frequency correlation with phenanthroline pK_a values in mixed copper complexes. Inorg Chim Acta 288(1):106–111
59. Tao L, Han J, Fu-Ming T (2008) Correlations and predictions of carboxylic acid pK_a values using intermolecular structure and properties of hydrogen-bonded complexes. J Phys Chem A 112(4):775–782
60. Psciuk BT, Prémont-Schwarz M, Koeppe B, Keinan S, Xiao D, Nibbering ETJ, Batista VS (2015) Correlating photoacidity to hydrogen-bond structure by using the local O-H stretching probe in hydrogen-bonded complexes of aromatic alcohols. J Phys Chem A 119(20):4800–4812
61. Sakti AW, Nishimura Y, Nakai H (2018) Rigorous pK_a estimation of amine species using density-functional tight-binding-based metadynamics simulations. J Chem Theory Comput 14(1):351–356
62. Oszczapowicz J (1991) Basicity, H-bonding, Tautomerism and Complex Formation of Imidic Acid Derivatives, Volume 2 edn. Wiley, Ltd., Chichester, Chapter 12, pp 623–688
63. Huang Y, Liu L, Liu W, Liu S, Liu S (2011) Modeling molecular acidity with electronic properties and Hammett constants for substituted benzoic acids. J Phys Chem A 115:14697–14707
64. Tomsho JW, Pal A, Hall DG, Benkovic SJ (2012) Ring structure and aromatic substituent effects on the pK_a of the benzoxaborole pharmacophore. ACS Med Chem Lett 3:48–52
65. Kreye WC, Seybold PG (2001) Correlations between quantum chemical indices and the pK_a's of a diverse set of organic phenols. Int J Quant Chem 109(15):3679–3684
66. Tao Y, Zou W, Cremer D, Kraka E (2017) Characterizing chemical similarity with vibrational spectroscopy: New insights into the substituent effects in monosubstituted benzenes. J Phys Chem A 121(42):8086–8096
67. Tao Y, Zou W, Cremer D, Kraka E (2017) Correlating the vibrational spectra of structurally related molecules: A spectroscopic measure of similarity. J Comput Chem 39(6):293–306
68. Konkoli Z, Cremer D (1998) A new way of analyzing vibrational spectra. I. Derivation of adiabatic internal modes. Int J Quantum Chem 67(1):1–9
69. Wilson EB, Decius JC, Cross PC (1955) Molecular Vibrations. McGraw-Hill, New York
70. Wilson JrEB (1939) A method of obtaining the expanded secular equation for the vibration frequencies of a molecule. J Chem Phys 7:1047
71. Kraka E, Freindorf M, Cremer D (2013) Chiral discrimination by vibrational spectroscopy utilizing local modes. Chirality 25(3):185–196
72. Zhang X, Dai H, Yan H, Zou W, Cremer D (2016b) B-H$\cdots\pi$ interaction: A new type of nonclassical hydrogen bonding. J Am Chem Soc 138(13):4334–4337
73. Kalescky R, Kraka E, Cremer D (2013) Local vibrational modes of the formic acid dimer—the strength of the double hydrogen bond. Mol Phys 111(9-11):1497–1510
74. Kalescky R, Zou W, Kraka E, Cremer D (2012) Local vibrational modes of the water dimer—comparison of theory and experiment. Chem Phys Lett 554:243–247
75. Tao Y, Zou W, Jia J, Li W, Cremer D (2017) Different ways of hydrogen bonding in water—why does warm water freeze faster than cold water? J Chem Theory Comput 13(1):55–76
76. Freindorf M, Kraka E, Cremer D (2012) A comprehensive analysis of hydrogen bond interactions based on local vibrational modes. Int J Quantum Chem 112(19):3174–3187
77. Oliveira V, Kraka E, Cremer D (2016) The intrinsic strength of the halogen bond: Electrostatic and covalent contributions described by coupled cluster theory. Phys Chem Chem Phys 18(48):33031–33046
78. Oliveira V, Kraka E, Cremer D (2017) Quantitative assessment of halogen bonding utilizing vibrational spectroscopy. Inorg Chem 56(1):488–502
79. Oliveira V, Cremer D (2017) Transition from metal-ligand bonding to halogen bonding involving a metal as halogen acceptor a study of Cu, Ag, Au, Pt, and Hg complexes. Chem Phys Lett 681:56–63
80. Setiawan D, Kraka E, Cremer D (2015) Strength of the pnicogen bond in complexes involving group Va Elements N, P, and As. J Phys Chem A 119(9):1642–1656
81. Setiawan D, Kraka E, Cremer D (2014) Description of pnicogen bonding with the help of vibrational spectroscopy—the missing link between theory and experiment. Chem Phys Lett 614:136–142
82. Setiawan D, Cremer D (2016) Super-pnicogen bonding in the radical anion of the fluorophosphine dimer. Chem Phys Lett 662:182–187
83. Oliveira V, Cremer D, Kraka E (2017) The many facets of chalcogen bonding: Described by vibrational spectroscopy. J Phys Chem A 121(36):6845–6862
84. Sethio D, Oliveira V, Kraka E (2018) Quantitative assessment of tetrel bonding utilizing vibrational spectroscopy. Molecules 23(11):2763
85. Kalescky R, Kraka E, Cremer D (2014) Description of aromaticity with the help of vibrational spectroscopy: Anthracene and phenanthrene. J Phys Chem A 118(1):223–237
86. Setiawan D, Kraka E, Cremer D (2016) Quantitative assessment of aromaticity and antiaromaticity utilizing vibrational spectroscopy. J Org Chem 81(20):9669–9686
87. Li Y, Oliveira V, Tang C, Cremer D, Liu C, Ma J (2017) The peculiar role of the Au3 unit in Au_m clusters: σ-aromaticity of the $Au5Zn^+$ ion. Inorg Chem 56(10):5793–5803
88. Cremer D, Kraka E (2017) Generalization of the tolman electronic parameter: The metal-ligand electronic parameter and

the intrinsic strength of the metal-ligand bond. Dalton Transac 46(26):8323–8338
89. Kalescky R, Kraka E, Cremer D (2014) New approach to Tolman's electronic parameter based on local vibrational modes. Inorg Chem 53(1):478–495
90. Setiawan D, Kalescky R, Kraka E, Cremer D (2016) Direct measure of metal–ligand bonding replacing the tolman electronic parameter. Inorg Chem 55(5):2332–2344
91. Kraka E, Larsson JA, Cremer D (2010). In: Grunenberg J (ed) Computational Spectroscopy: Methods, Experiments and Applications. Wiley, New York
92. Kalescky R, Kraka E, Cremer D (2013) Identification of the strongest bonds in chemistry. J Phys Chem A 117(36):8981–8995
93. Kraka E, Setiawan D, Cremer D (2015) Re-evaluation of the bond length-bond strength rule: the stronger bond is not always the shorter bond. J Comp Chem 37(1):130–142
94. Setiawan D, Kraka E, Cremer D (2015) Hidden bond anomalies: the peculiar case of the fluorinated amine chalcogenides. J Phys Chem A 119(36):9541–9556
95. Humason A, Zou W, Cremer D (2014) 11,11-dimethyl-1,6-methano[10]annulene-an annulene with an ultralong CC bond or a fluxional molecule? J Phys Chem A 119(9):1666–1682
96. Kalescky R, Zou W, Kraka E, Cremer D (2014) Quantitative assessment of the multiplicity of carbon–halogen bonds: carbenium and halonium ions with F, Cl, Br, and I. J Phys Chem A 118(10):1948–1963
97. Kraka E, Cremer D (2009) Characterization of CF bonds with multiple-bond character: bond lengths, stretching force constants, and bond dissociation energies. ChemPhysChem 10(4):686–698
98. Chai JD, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. Chem Phys 10:6615
99. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JAJr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian16 Revision A.03. Gaussian Inc., Wallingford
100. Dennington R, Keith TA, Millam JM (2016) Gaussview Version 5.0.9. Semichem Inc., Shawnee Mission
101. Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Phys Chem 90:1007
102. Woon DE, Dunning TH (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. J Phys Chem 98:1358
103. Wilson AK, Woon DE, Peterson KA, Dunning TH (1999) Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. J Phys Chem 110:7667
104. Hehre WJ, Ditchfield R, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. J Chem Phys 56:2257
105. Francl MM, Petro WJ, Hehre WJ (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. J Chem Phys 77:3654
106. Rassolov V, Pople JA, Ratner M, Windus T (1998) 6-31G* basis set for atoms K through Zn. J Chem Phys 109:2257
107. Ali ST, Karamat S, Kóna J, Fabian WMF (2010) Theoretical prediction of pK_a values of seleninic, selenenic, sulfinic, and carboxylic acids by quantum-chemical methods. J Phys Chem A 114(47):12470–12478
108. Kraka E, Zou W, Filatov M, Tao Y, Grafenstein J, Izotov D, Gauss J, He Y, Wu A, Konkoli Z (2017) COLOGNE2017. See http://www.smu.edu/catco
109. Helgaker T, Coriani S, Jorgensen P, Kristensen K, Olsen J, Ruud K (2012) Recent advances in wave function-based methods of molecular-property calculations. Chem Rev 112(1):543–631
110. D'Alessandro DM, Smit B, Long JR (2010) Carbon dioxide capture: Prospects for new materials. Angew Chem Int Ed 49:6058
111. MacDowell N, Florin N, Buchard A, Hallett J, Galindo A, Jackson G, Adjiman CS, Williams CK, Shah N, Fennell P (2010) An overview of CO_2 capture technologies. Energy Environ Sci 3:1645
112. Kenarsari SD, Yang D, Jiang G, Zhang S, Wang J, Russell AG, Wei Q, Fan M (2013) Review of recent advances in carbon dioxide separation and capture. RSC Adv 3:22739
113. Yang X, Rees RJ, Conway W, Puxty G, Yang Q, Winkler DA (2017) Computational modeling and simulation of CO_2 capture by aqueous amines. Chem Rev 117:9524–9593
114. Rochelle GT (2009) Amine scrubbing for CO_2 capture. Science 325:1652
115. Mumford KA, Wu Y, Smith KH, Stevens GW (2015) Review of solvent based carbon-dioxide capture technologies. Front Chem Sci Eng 9:125
116. Versteeg GF, van Dijck LAJ, van Swaaij WPM (1996) On the kinetics between CO_2 and alkanolamines both in aqueous and non-aqueous solutions. An overview. Chem Eng Commun 144:113
117. Puxty G, Rowland R, Allport A, Yang Q, Bown M, Burns R, Maeder M, Attalla M (2009) Carbon dioxide postcombustion capture: A novel screening study of the carbon dioxide absorption performance of 76 amines. Environ Sci Technol 43:6427
118. Rayer AV, Sumon KZ, Jaffari L, Henni A (2014) Dissociation constants (pK_a) of tertiary and cyclic amines: Structural and temperature dependences. J Chem Eng Data 59:3805
119. Tagiuri A, Mohamedali M, Henni A (2016) Dissociation constant (pK_a) and thermodynamic properties of some tertiary and cyclic amines from (298 to 333) K. J Chem Eng Data 61:247
120. Hamborg ES, Versteeg GF (2009) Dissociation constants and thermodynamic properties of amines and alkanolamines from (293 to 353) K. J Chem Eng Data 54:1318
121. Manallack DT (2007) The pK_a distribution of drugs: Application to drug discovery. Perspectives in Medicinal Chemistry 1:25–38
122. Charifson PS, Walters WP (2014) Acidic and basic drugs in medicinal chemistry: a perspective. J Med Chem 57:9701–9717
123. Mitani GM, Steinberg I, Lien EJ, Harrison EC, Elkayam U (1987) The pharmacokinetics of antiarrhythmic agents in pregnancy and lactation. Clin Pharmacokinet 12(4):253–291
124. Xie X, Steiner SH, Bickel MH (1991) Kinetics of distribution and adipose tissue storage as a function of lipophilicity and chemical structure. II Benzodiazepines. Drug Metab Dispos: The Biological Fate of Chemicals 19(1):15–9
125. Avdeef A (2001) Physicochemical profiling (solubility, permeability and charge state). Curr Top Med Chem 1(4):277–351
126. Kerns E, Di L (2004) Physicochemical profiling: Overview of the screens. Drug Discov. Today Technol. 1(4):343–348