Theoretical insights into the linear relationship between $pK_a$ values and vibrational frequencies

Mateus Quintano, Elfi Kraka*

Computational and Theoretical Chemistry Group (CATCO) Department of Chemistry, Southern Methodist University Dallas, TX 75275, USA

**ARTICLE INFO**

**Keywords:**
Ansatz-based equation
$pK_a$
Vibrational spectroscopy
Linear correlation
Characterization of normal modes

**ABSTRACT**

A linear relationship between $pK_a$ values and vibrational frequencies has been derived in this work for hydrogen halides based on a statistical thermodynamic ansatz and successfully tested for a set of 24 aliphatic carboxylic acids, with predicted $pK_a$ values differing from measured values by less than 0.3 $pK_a$ units. Our work opens the door to reliably predicting $pK_a$ values for families of compounds often difficult to measure and paves the way for a natural extension of the derived relationship referring to local vibrational mode theory to treat systems with more delocalized A-H stretching vibrations, such as nitroacetic acid.

1. Introduction

The $pK_a$ is defined as the negative logarithm (to the base of 10) of the acid dissociation constant ($K_a$) for the dissociation reaction HA $\rightleftharpoons$ H$^+$ + A$^-$ of the acid HA:

$$pK_a = -\log K_a$$

where $K_a$ is given by

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

A diverse number of computational approaches for the calculation of $pK_a$ estimates have been suggested over time, some representative examples are collected in Refs. [1–18]. However, there is still a lot of room for improvement; an efficient but also reliable protocol for the determination of $pK_a$ (i.e., errors in range of 0.5 - 1 $pK_a$ units or less) is expected to find application in various fields, including drug design [3,8,19,11] and crystal engineering [20–26]. One popular approach is to determine the dissociation free energy via a thermodynamic cycle [17,18] to be used in a linear free energy relationship (LFER) to estimate $pK_a$ values [4,7,16]. However, the total acidity resulting from the entire molecule, rather than the acidity of a particular functional group, is accounted for. On the other hand, the normal vibrational A-H stretching frequency of the acid HA has been suggested as a more local acidity descriptor, coined as the so-called *hydrogen-bonding probe method* [12–15]. Using this approach, e.g., molecular acidity as reflected by the $pK_a$ has been reported to correlate linearly with the O-H stretching frequency for selected organic acids. Interestingly, both LFER and hydrogen-bonding probe methods involve linear equations in which two coefficients are fitted to experimental $pK_a$ values and the corresponding calculated properties, free energy and vibrational frequency, respectively.

In this work we explored if the observed linear relationship based on vibrational frequencies can be generalized and furnished with a solid physical foundation. The first part of the investigation sets about tackling the problem in the context of small acid molecules whose vibrational degree of freedom can be fairly approached by a gas-phase oscillator, namely, the simple set of hydrogen halides, i.e., HF, HCl, HBr, and HI. We utilized an ansatz based on the canonical ensemble of statistical thermodynamics in the quantum mechanical formulation [27]. Throughout the development of our ideas, we found useful guidance in Professor Kuchitsu’s work on vibrational spectroscopy [28–31]. After deriving a proof of concept, the second part of the study then discusses the occurrence of the linear relationship for the set of 24 aliphatic carboxylic acids, shown in Fig. 1, adapted from Ref. [13]. The family of carboxylic acids was chosen as a cost-effective and opportunate example to illustrate, in a solvation-improved version in comparison with the literature, the importance of the $pK_a$ probe to capture the linear relationship for local acidity in polyatomic systems where steric hindrance and substituent effects are present [13].

An important prerequisite for this relationship is that the normal stretching mode of the associated normal mode A-H stretching frequency is of A-H stretching character, i.e., it is localized in the A-H bond...
and is not involved in mode-mode coupling with other modes, which is often found for molecular systems [32]. This important prerequisite forms the basis of the hydrogen-bonding probe method, but to our best knowledge, never has been carefully inspected. We used in this work as an assessment tool the characterization of normal vibrational mode (CNM) procedure, which is part of the local mode analysis (LMA) developed in our group [33–37]. A comprehensive review of LMA can be found in Ref. [37]. CNM decomposes normal vibrational modes into local mode contributions and has led to a new way of analyzing vibrational spectra, which has been successfully applied to assess the validity of so-called Stark effect probes relying on a localized bond stretching probe [38] or help assigning absorption bands in benziodoxoles and benziodazolotetrazoles, where the latter are well known for their high positive enthalpy of formation, making them attractive as propellants, explosives, and energetic materials [39], just to name two examples.

2. Methodology

For the purpose of assessing our new ansatz, reported pK_a values in aqueous solution for HCl, HBr, and HI, combining experimental and computational approaches [9], as well as the tabulated pK_a value in aqueous solution for HF [40], were utilized in addition to experimental fundamental frequencies for the diatomics, HF, HCl, HBr, and HI [41].

Geometry optimization and subsequent normal vibrational frequency calculations, followed by CNM for the set of 24 aliphatic carboxylic acids, were performed with Gaussian 16 [42] utilizing the ωB97X-D density functional [43] in combination with Dunning’s aug-cc-pVTZ basis set [44,45]. Levels 3-6 involved both geometry optimization and vibrational frequency calculation of the hydrogen-bonded complexes (acid and probe). An ultra-fine grid integration [46] and a tight convergence criterion was applied for the SCF procedure. CNM was
performed with the standalone LModeA package [47]. The combination of the ωB97X-D functional with the aug-cc-pVTZ basis set is justified by Refs. [5,1]. Optimized geometries of all systems but that of cis-formic acid at Level 6 were in trans conformation.

Fig. 2 sketches different scenarios which were tested in this work; gas phase (Level 1), aqueous solution modeled by the Polarizable Continuum Model (PCM) [48] (Level 2), aqueous solution with one ammonia molecule attached (Level 3); gas phase with one water molecule attached (Level 4), gas phase with one ammonia molecule attached (Level 5) and aqueous solution with one water molecule attached (Level 6), which yielded the best linear correlation between \( pK_a \) and \( \bar{v} \), and therefore is discussed herein. Additional results for Levels 1-5 are collected in the Supplementary Material. These results are in line with the recent work of Dutra et al. who reported that, in addition to the PCM dielectric medium, the explicit addition of solvent molecules tends to improve results [1]. An important point was to reevaluate for carboxylic acids 1–24 the "ammonia probe" (Level 5 in Fig. 2) suggested in the literature [13], and to explore the performance of the "water probe" for this set of 24 molecules including a different combination of functional and basis set (see Supplementary Material). In this way, it could be systematically tested if previously published results were biased by the use of the "ammonia probe" [13]. Furthermore, \( pK_a \) values are generally reported for aqueous solution, therefore the "water probe" seems to be a more natural choice as e.g., for the tabulated \( pK_a \) values in aqueous solution which were plotted against the computed normal vibrational frequencies associated with O-H stretching vibrations [40]. In summary, the simple linear regression with two-dimensional sample points served the purpose of testing our hypotheses by linear fitting in comparison with previously published results [13].

3. Results and Discussion

3.1. Ansatz for Hydrogen Halides

In aqueous solution, the concentration of undissociated strong acids is incomparably smaller than that of the dissociated species. This is the case for HI, HBr and HCl, as reflected by their recommended \( pK_a \) values [9]. For HF, the tabulated \( pK_a \) value refers to a different combination of functional and basis set (see Supplementary Material). In this way, it could be systematically tested if previously published results were biased by the use of the "ammonia probe" [13]. Furthermore, \( pK_a \) values are generally reported for aqueous solution, therefore the "water probe" seems to be a more natural choice as e.g., for the tabulated \( pK_a \) values in aqueous solution which were plotted against the computed normal vibrational frequencies associated with O-H stretching vibrations [40]. In summary, the simple linear regression with two-dimensional sample points served the purpose of testing our hypotheses by linear fitting in comparison with previously published results [13].

![Fig. 2. Schematic of levels 1-6 (from worst to best linear \( pK_a \) versus \( \bar{v} \) correlations) for the reassessment of "water" and "ammonia probes", employing the ωB97X-D functional with solvation effects (PCM) in conjunction with the aug-cc-pVTZ basis set. The PCM dielectric medium is indicated below each general schematic representation for the test examples 1–24; "none" corresponds to gas phase calculations.](image)

![Fig. 3. Correlation between experimental \( pK_a \) values and fundamental vibrational frequencies for the family of hydrogen halides. From the linear regression, predicted \( pK_a \) values can be calculated: \( pK_a(\text{pred}) = -27.67 + 0.007708\bar{v} \). Recommended \( pK_a \) values were taken from the literature, see text.](image)
which can be plugged into Eq. (5) to give

\[ -\log K_a = \gamma + \frac{\log(e)c}{2k_BT} \left( \frac{1}{2x} - 1 \right) \tilde{v} \]  

\[ (8) \]

Combining Eqs. (8) and (1) by hypothesis yields

\[ pK_a = \gamma - \frac{\log(e)c}{2k_BT} \left( \frac{1}{2x} - 1 \right) \tilde{v} \]  

\[ (9) \]

Therefore, our ansatz makes the following predictions for hydrogen halides:

- Within the acid family, if the term \( \gamma \) is assumed to be system-independent and the anharmonicity constant is taken as an approximately constant value, the resulting \((\tilde{v}, pK_a)\) curve is expected.
- The major contribution to the slope of the resulting \((\tilde{v}, pK_a)\) curve originates from \( D_\tilde{v} \) as 1/2xa=1 in Eq. (9). To put it another way, the slope is controlled by an electronic effect.
- For relatively large anharmonicity, which is the case for the hydrogen halide family (x \( \approx \) 0.02) [41], the slope is anticipated to be in the order of \( 10^{-2} \) cm at the absolute temperature of 298.15 K.

The examination of the above hypotheses is illustrated in Fig. 3. Although visible uncertainty comes from the recommended \( pK_a \) value of HI [9], the coefficient of determination \((R^2)\) of the simple linear regression still reflects the satisfactory replication of outcomes by our model function (see Supplementary Material for details). Indeed, there is a linear relationship our ansatz predicts. The slope of approximately 0.01 cm is in harmony with the order of magnitude anticipated by Eq. (9) at 298.15 K. Our theoretical framework is an approximation since \( pK_a \) values are measured in solution, whereas Eq. (9) was explicitly derived for the oscillator of a diatomic molecule in the gaseous phase. Even so, it provides a much better understanding of the linearity that arises from the relationship between the \( pK_a \) and the vibrational frequency for hydrogen halides. Furthermore, our concept can be utilized to gain insights into and to inspect similar relationships between \( pK_a \) and A-H normal vibrational frequencies reported in the literature such as in the work of Tao et al. [15,14,13,12]. Combined with CNM it can be carefully checked if in the case of larger acids e.g., with more than one polar functional group, the prerequisite of "local acidity" described

Fig. 4. Correlation between \( pK_a \) values and normal mode O-H stretching frequencies \((\tilde{v})\) for aliphatic carboxylic acids 1–24. Level 6 calculations: \( \alpha \)B97X-D (PCM)/aug-cc-pVTZ level of theory, with an explicit water molecule, in the cavity within the water dielectric medium.
above is still valid. This is demonstrated in the following for a series of 24 aliphatic carboxylic acids, reported by Tao et al. [13].

### 3.2. Aliphatic Carboxylic Acids

We found a significant linear correlation between \( \tilde{\nu} \) and \( pK_a \) (\( R^2=0.9926 \)) for the 24 aliphatic carboxylic acids 1–24 investigated in this work, as depicted in Fig. 4. It has to be noted that this strong correlation was achieved by the inclusion of solvation effects via PCM combined with an explicit water molecule (see Level 6 in Fig. 2). As shown in Table 1, the linear regression takes on the form of Eq. (9), with the slope in the order of \( 10^{-2} \) cm \( (pK_a \text{ (pred)} = -41.018 + 0.013516 \tilde{\nu}) \).

The predicted \( pK_a \) values exhibit an excellent agreement with the experimental data with average differences between experimental and predicted \( pK_a \) values being smaller than 0.3 \( pK_a \) units.

As an assessment tool for the local character of the normal O-H stretching vibration in compounds 1–24, CNM was applied. Fig. 5 summarizes the results by displaying for each of the 24 carboxylic acids the % of local O-H stretching mode character. As obvious from Fig. 5, the normal modes associated with the O-H stretching vibrations are more than 94% localized. Such a high localization strongly aligns with the family of hydrogen halide diatomics, in which the only normal vibrational mode is unmistakably local, and it also demonstrates the sensitivity of \( pK_a \) values to local acidity. However, there is one exception, nitroacetic acid 21. As depicted by Table 1 and Fig. 5, there are two normal vibrations in the typical O-H vibrational stretching range, 3335 and 3338 cm\(^{-1}\) (only 3 cm\(^{-1}\) apart) with a local O-H stretching mode character of only 52% and 42%, respectively.

In Fig. 6 the O-H normal vibrational modes associated with the normal O-H stretching vibration in nitroacetic acid 21 are shown. As revealed in Fig. 6a, in the gas phase the normal O-H stretching vibration involves only the O and H atoms of the OH group, i.e., it is perfectly localized leading to a 100% local O-H contribution to this mode. However, in aqueous solution a different picture emerges, as illustrated by Figs. 6b and 6c. There is a coupling between the local O-H stretching mode and the local C-H stretching modes of the CH\(_2\) group. In the normal mode associated with the 3135 cm\(^{-1}\) normal mode vibrational frequency, C-H and O-H stretching motions are coupled in a symmetric fashion (i.e., the O-H and the C-H vibrations move in the same direction, see Fig. 6b) with ca. 23% C-H stretching contribution to this normal mode. In the normal mode associated with the 3138 cm\(^{-1}\) vibrational frequency, C-H and O-H stretching motions are coupled in an asymmetric fashion (i.e., the O-H and the C-H vibrations move in opposite directions, see Fig. 6c) with ca. 28% C-H stretching contribution to this normal mode. Although the predicted \( pK_a \) value for nitroacetic acid 21 deviates only by -0.12 \( pK_a \) units from the experimental value, the CNM analysis indisputably clarifies that the prerequisite of a localized O-H vibration is not given and therefore, the good agreement between

![Fig. 5. Decomposition of the normal mode associated with the O-H stretching vibration into local mode contributions for aliphatic carboxylic acids 1–24. The second normal O-H frequency exhibited by 21 (3138 cm\(^{-1}\)) is given by 21'. Level 6 calculations: \( \omega B97X-D(PCM)/aug-cc-pVTZ \) level of theory, with an explicit water molecule, in the cavity within the water dielectric medium.](image)

![Fig. 6. Visualization of O-H normal vibrational mode associated with the normal O-H stretching vibration in nitroacetic acid 21; the movement of atoms involved in the vibration is indicated by arrows. (a) Level 1 calculation: \( \omega B97X-D/aug-cc-pVTZ \) level of theory (gas phase); \( \tilde{\nu} = 3818 \) cm\(^{-1}\), 100% local O-H stretching character; (b) and (c) Level 6 calculations: \( \omega B97X-D(PCM)/aug-cc-pVTZ \) level of theory (explicit water molecule, in the cavity within the water dielectric medium); (b) \( \tilde{\nu} = 3135 \) cm\(^{-1}\), 52% local O-H stretching character; and (c) \( \tilde{\nu} = 3338 \) cm\(^{-1}\), 42% local O-H stretching character.](image)
experiment and prediction could be just by coincidence, resulting from the coupling between stretching motions with comparable vibrational frequencies.

4. Conclusion

We have devised a statistical thermodynamic ansatz on the basis of the canonical ensemble equations in the quantum mechanical form. Our ansatz successfully predicts the linear relationship between the pK_a and the fundamental vibrational frequency for the family of hydrogen halides. The linearity extends to larger systems, such as the family of aliphatic carboxylic acids, with predicted pK_a values differing form measured values by less than 0.3 pK_a units. As demonstrated in this work, for nitroacetic acid 21, CNM is a powerful tool to assess the underlying prerequisite of a localized A-H stretching vibration which may no longer hold in solution for larger systems, leading to extensive mode-mode coupling. The inclusion of solvent effects via PCM, extended by an explicit water molecule, led to the best correlation in reassessing the hydrogen-bonding probe method for a series of 24 carboxylic acids. In summary, our work opens the door to reliably predicting pK_a values for families of compounds often difficult to measure and pave the way for a natural extension of the derived relationship, by referring to local vibrational A-H modes instead of A-H normal modes for systems with more delocalized A-H stretching vibrations, such as nitroacetic acid, or that demand a sound bond strength descriptor. Work is in progress in this direction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank SMU’s Center for Scientific Computing for providing generous computational resources. This work was financially supported by the National Science Foundation, Grant CHE 2102461.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cplett.2022.139746.

References


