



The AIBLHiCoS Method: Predicting Aqueous pK_a Values from Gas-Phase Equilibrium Bond Lengths

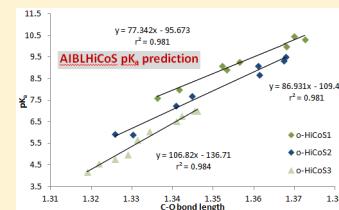
Cate Anstöter,^{†,‡} Beth A. Caine,^{†,‡} and Paul L. A. Popelier*,^{†,‡}

[†]Manchester Institute of Biotechnology (MIB), 131 Princess Street, Manchester M1 7DN, United Kingdom

[‡]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom

Supporting Information

ABSTRACT: The proposed AIBLHiCoS method predicts a given compound's pK_a in aqueous solution from a single ab initio bond length only, after geometry optimization in the gas phase. Here we provide simple and predictive equations for naphthols and chemically similar biomolecules. Each linear equation corresponds to a *High-Correlation Subset* (HiCoS) that expresses the novel type of linear free energy relationship discovered here. The naphthal family exhibits a clear and strong relationship with the phenol family, with the "active" C–O bond always producing the highest correlations. The proposed method can isolate erroneous experiments and operate in non-aqueous solution and at different temperatures. Moreover, the existence of "active fragments" is demonstrated in a variety of sizable biomolecules for which the pK_a is successfully predicted.



1. INTRODUCTION

One of the most important and powerful quantitative parameters in physical organic chemistry is the dissociation constant, commonly expressed as pK_a , for any weakly acidic or basic groups (i.e., between 2 and 11 pK_a units) in small molecules or macromolecules. The comparison of pK_a values across chemical functionalities was one of the first historical successes in the quantification of chemical reactivity, thereby giving mechanistic insight. An understanding of the ability of a compound to donate or accept a proton is fundamental to understanding both chemical and biological processes and has seen the publication of models that account for and predict new chemical affinities.

Computational models that can accurately predict pK_a values have been an area of much interest in the theoretical chemistry community for several decades and continuing to date. Very recently, Ugur et al.¹ distinguished four theoretical methods that can estimate pK_a values in proteins: (i) knowledge-based empirical methods that utilize local and global environments to predict the pK_a values;^{2–4} (ii) molecular-modeling-based methods, which modify ionizable sites to model single or multiple protonation or unprotonation states;^{5–9} (iii) quantum-based methods, which typically use thermodynamic cycles in pK_a prediction;^{10–16} and (iv) methods based on the polarizable continuum solvent model, which rely on the dielectric difference between the protein interior and the external solvent.^{17–20}

The most popular of the accurate methods are based on a thermodynamic cycle in which the free energy of the reaction in solution and the free energy of the corresponding reaction in vacuo are resolved. By the use of continuum representations of the solvent, the difference in the solvation free energies of the products and reactants is computed. The absolute free energy for the protonation is then converted to a pK_a via the absolute free energy of solvation of the proton. Main areas in which

differences in these models arise are in the thermodynamic cycle on which the theoretical calculation is based, the level of theory used for the gas-phase calculation, and the solvation model used. Implicit-solvent models often have trouble correctly describing solvent effects, which is a critical step in accurately predicting pK_a by computation.

Methods based on quantitative structure–activity relationships are also utilized for in silico pK_a modeling because of their predictive and diagnostic abilities. For a very recent example, collaboration between Bayer Pharma and Simulations Plus, Inc. has produced a prediction tool of outstanding quality.²¹ Their paper introduces and demonstrates the importance of understanding the difference between macrostate and microstate pK_a values for molecules with multiple ionizable groups. The authors highlighted the lack of distinction between these two parameters as a main shortfall of pK_a prediction packages based on quantitative structure–property relationships (QSPRs). In contrast to macroconstants, which are commonly used in previous QSPR models, microconstants are properties of individual groups and depend directly on the structure of the microstate involved, making them natural candidates for use in QSPR modeling.²¹

The method used in the present paper is called AIBLHiCoS for reasons that will become clear below. AIBLHiCoS was previously developed by this group; it is remarkably simple and has produced accurate results over rigorous testing in previous work.^{22–26} AIBLHiCoS exploits a novel linear free energy relationship (LFER) that connects a single "active" bond length (in gas-phase ab initio equilibrium geometries) with the experimental pK_a values in aqueous solution of a family of compounds. In previous publications this LFER has been successfully demonstrated on a series of substituted phenols,

Received: September 22, 2015

Published: January 28, 2016



benzoic acid and anilines, guanidines, bicyclo[2.2.2]octane, and cubane carboxylic acids. This novel model has allowed prediction of pK_a values of even notoriously difficult functional groups, such as guanidine,²⁵ with accuracies that outperform those of other commercial software tools. Exploration of the correlation between the chemically active bond and the experimental pK_a of an acidic proton in close proximity has led to an understanding that chemical space can be partitioned into *high-correlation subsets* (HiCoSs). HiCoSs contain molecules that have structural or chemically commonality that is dictated by the LFER of pK_a versus bond length itself. The origin of the acronym AIBLHiCoS (Ab Initio Bond Length High Correlation Subsets) is now clear.

Systematic recognition of HiCoSs within a data set leads to a drastic improvement of the r^2 correlation coefficient between bond length and experimental pK_a , with typical values of 0.9 or above. Through identification of the HiCoS by which a new molecule would best be represented, the predicted pK_a can be found through calculation of the ab initio bond length and insertion of this value into the equation corresponding to the HiCoS. Any outliers are rationalized in terms of the commonality, or lack thereof, that they share with other compounds in a HiCoS. Establishing a high correlation for the right physicochemical reason is at the heart of the AIBLHiCoS method. As a result, any outlier presented in this and previous work is confidently reassigned to another HiCoS in which the chemical commonality is higher.

This work focuses on the prediction of pK_a for the naphthol family, for which there exists a wealth of information in the literature. Previous experimental and theoretical work has studied the effect of substituents on the pK_a values of both 1- and 2-naphthols²⁷ and the pK_a^* or photoacidity of the naphthols.^{28–32} Naphthol is incorporated in dyes, in medicinal compounds used as bacterial antibiotics, estrogenic steroids, and chemotherapy medications, and in photodynamic therapy for cancer treatment as a fragment of a larger molecule. The prediction of pK_a is of particular interest in drug design and structural biology. Knowledge of the pK_a values for ionizable groups is essential in drug design in order to modulate physical and pharmacological properties such as distribution, bioavailability, and solubility.⁷ The limitations of X-ray crystallography, due to the very low electron density of hydrogen atoms, leave ambiguity in the unresolved three-dimensional positions of protons in a protein.³³ The dependence on the knowledge of the protonation states of residues to understand enzyme catalysis mechanistically drives the need for the development of an accurate pK_a predictor that can provide experimentally unavailable parameters.

2. COMPUTATIONAL METHODS

Geometries were optimized at a number of levels of theory with respect to both methods (e.g., HF, B3LYP, MP2) and basis sets. Our previous work^{22–26} showed that Hartree–Fock (HF) calculations with the modest 6-31G(d) basis set provided statistical models similar to those for larger basis sets. For the family of substituted naphthols studied here, this level of theory did not suffice, and a drastic improvement was seen in the required equilibrium bond lengths using density functional theory (DFT) with the B3LYP functional and the slightly larger 6-31G(d,p) basis set. Frequency calculations were carried out to confirm that the geometries obtained were indeed energetic minima. All of the calculations were executed using the Gaussian 09 program.

Optimizations using the MP2 method were deemed inappropriate in view of the previously documented poor representation of aromatic systems^{34,35} due to the prediction of nonplanar structures. Application of MP2/6-31G(d,p) to calculate the equilibrium geometries of the two reference species, 1-naphthol and 2-naphthol, indeed returned planar structures with a single imaginary frequency, each corresponding to a puckering of the rings, at 317*i* and 326*i* cm⁻¹, respectively. Discussion of this and similar discrepancies for aromatic systems can be found in refs 34 and 35. Other work³⁶ identified B3LYP/6-311G(d,p) as a benchmark level of theory for calculations of physical properties of aromatic systems. All of the geometries investigated within the current work were optimized using this level of theory. However, no significant improvement was noted for any of the correlations, with decreases in correlation seen for all HiCoSs.

Figure S1 in the Supporting Information (SI) compares the seven levels of theory investigated for calculation of the C–O bond lengths for the *meta* and *para* HiCoS (denoted as “m+p-HiCoS”), with any common outliers removed. Table S1 in the SI gives the 19 substituents of this HiCoS, while Table S2 provides the r^2 values for all seven levels of theory. The root-mean-square errors of estimation (RMSEEs) and cross-validation statistics in the form of leave-one-seventh-out q^2 values were obtained using SIMCA-P; details of these statistical techniques have been covered in depth in a previous publication.²³ The best level of theory was found to be B3LYP/6-31G(d,p) ($r^2 = 0.949$), and overall B3LYP does consistently better than HF. A number of further observations were made. From Figure S1 it is clear that the range of C–O bond lengths was larger when HF was used for geometry optimizations. This is also reflected in the relatively low r^2 values for the HF levels of theory versus the B3LYP levels of theory, as shown in Table S2. For the B3LYP levels, the addition of a diffuse function to the basis set generated new outliers (and hence a deterioration of r^2). However, removal of the two most pronounced outliers, shown as circled points in Figure S1, returned substantially improved correlations of 0.877 and 0.872 for B3LYP/6-31+G(d) and B3LYP/6-31+G(d,p), respectively.

The van't Hoff equation was used to predict the change in pK_a for compounds that did not have experimental data available at 25 °C. The calculations were carried out in two stages: first, the enthalpy change for dissociation was calculated for compounds considered to be outliers (32 and 36 in Table S6 and 15 and 17 in Table S9). This quantity was estimated using B3LYP/6-311++G(d,p) for energy calculations, in line with previous studies on bond dissociation energies for phenols.³⁶ The formulas used for the enthalpy change and van't Hoff calculations can be found in the last section of the Supporting Information.

The data sets used contained compounds with multiple ionizable groups, some of which had a microscopic dissociation value lower than that of the acidic naphthol proton. In other words, a proton can dissociate first from elsewhere than the naphtholic O–H under study. In order to incorporate these compounds correctly, it is necessary to use ionic structures that represent this change in macrostate. Inclusion of ionic compounds dictates a change in the level of theory. The literature benchmark for gas-phase calculations incorporates a diffuse function in the basis set employed. However, in this work, this remedy was not sufficient and did not allow the deprotonated species to be considered as part of the HiCoS.

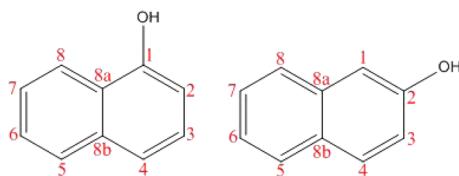
Calculations using the CPCM continuum model of solvation at the B3LYP/6-31G(d,p) level allowed successful incorporation of these charged species into the *meta* and *para* HiCoS with a reasonably high correlation ($r^2 = 0.88$).

3. RESULTS AND DISCUSSION

3.1. Phenols. The chemical family investigated in this paper is the naphthol family, a group of aromatic compounds that shares structural similarity to the phenol family. Naphthols are related to phenols through the addition of a second fused benzene ring. Because of the high level of structural similarity and the commonalities observed in the aromatic class of molecules, it was predicted that there would be an overlap with previous findings for bond lengths and pK_a in the phenol family by Harding and Popelier.²³ It was therefore expected that the same active bond, C–O, would provide a correlation with experimental pK_a values and that the same HiCoSs would emerge for the naphthol family, according to substituent proximity to the ipso carbon. An additional HiCoS is to be expected for the naphthols, accounting for the addition of the second ring to chemical space. The HiCoSs found for the naphthol family were the same as those established previously by Harding for the phenols: the data were split into a *meta* and *para* set (55 compounds) and an *ortho* set (90 compounds, of which 26 can form an internal hydrogen bond). The phenol *ortho* set was split into further HiCoSs, displaying the same conformational dependence as seen in the current work.

For both 1- and 2-substituted naphthols (Scheme 1, left and right panels, respectively), a large data set of experimental pK_a

Scheme 1. Two Isomers of Naphthol (1-Naphthol and 2-Naphthol), Showing the Labeling Scheme Used Throughout



values is available.^{27,31,37} When all of the data points were considered together, no bond length provided a good correlation. However, when the naphthols were considered by substituent proximity to the ipso carbon, the correlations clearly improved for preliminary HiCoSs. For all of the HiCoSs, the recurrent active bond is $r(\text{C}-\text{O})$. Table S3 gives the correlations between the experimental pK_a values and all of bond lengths around the primary ring (which contains the OH group associated with the pK_a). Table S3 shows that the correlation observed for the C–O bond ($r^2 = 0.95$) is strikingly high in comparison with those for the other bond lengths (which range from $r^2 = 0.01$ to $r^2 = 0.35$ only). The naphthol isomers were initially investigated separately, before HiCoSs with commonalities were combined.

The previous work on phenols was carried out using HF/6-31G(d), which is a different level than the one found to be necessary for the naphthols in the current study. Therefore, the *m*- and *p*-phenols were recalculated at the current level of theory for naphthols in order to provide a meaningful comparison. Initial work found the best correlations when only the nitro-, alkyl-, and halide-substituted *m*- and *p*-phenols were considered. Narrowing the *meta* and *para* data set down to include only the aforementioned functional groups gave

higher r^2 values, i.e., 0.87 for all *m*- and *p*-phenols as opposed to 0.97 for the narrowed set. When the calculations were redone at the level of theory used here for the naphthols, we obtained r^2 values of 0.91 for the full set of *m*- and *p*-phenols and 0.97 for the narrowed set. The improved correlation for the narrowed HiCoS of *m*- and *p*-phenols at the B3LYP/6-31G(d,p) level is shown in Figure 1 along with the corresponding *meta* and *para*

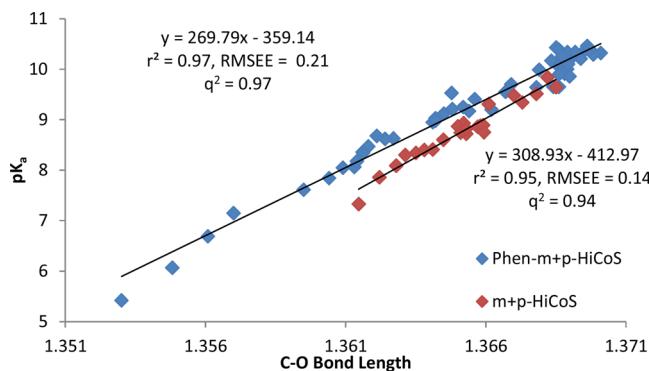
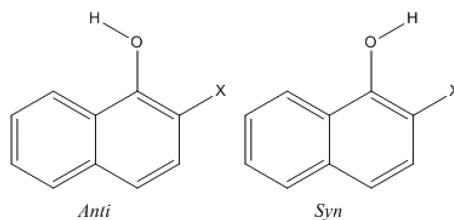


Figure 1. Correlations of ab initio C–O bond lengths with experimental pK_a values for the phenol (blue) and naphthol (red) families. The correlations shown are for the *meta* and *para* High Correlation Subsets (m+p-HiCoSs) identified for the two chemical families.

HiCoS found for the naphthol family. When the phenol and naphthol families are considered together, the correlations for the two are impressive, running almost parallel across a similar pK_a range, with naphthols showing longer C–O bonds than phenols for higher pK_a values across the range, indicative of heightened acidity.

3.2. Conformational Dependence. 3.2.1. Formation of Intramolecular Hydrogen Bonds. For the *ortho* subsets of both isomers, the conformation used affects the correlation between the experimental data and the gas-phase bond length, as has been seen in previous work.^{23,26} The proximity of the two substituents in either isomer's *ortho* subset causes a much greater perturbation to chemical space than in non-*ortho* subsets. Whether *ortho* substituents are *syn* or *anti* with respect to the alcohol group (Scheme 2) affects their ability to form

Scheme 2. Common Skeletons of the Two Conformers Formed by *o*-Naphthols That Are Capable of Forming Internal Hydrogen Bonds



intramolecular hydrogen bonds, which in turn affects the pK_a . A theoretical study³⁸ showed that for most *o*-phenols, the acidity of the OH group is reduced by the formation of intramolecular hydrogen bonds. However, a later study³⁹ suggested an increase in acidity upon formation of an intramolecular hydrogen bond. This result was noted in previous work by Harding: *o*-phenols adopting a conformation able to form an internal hydrogen bond had a lower pK_a , indicating heightened acidity. The latter

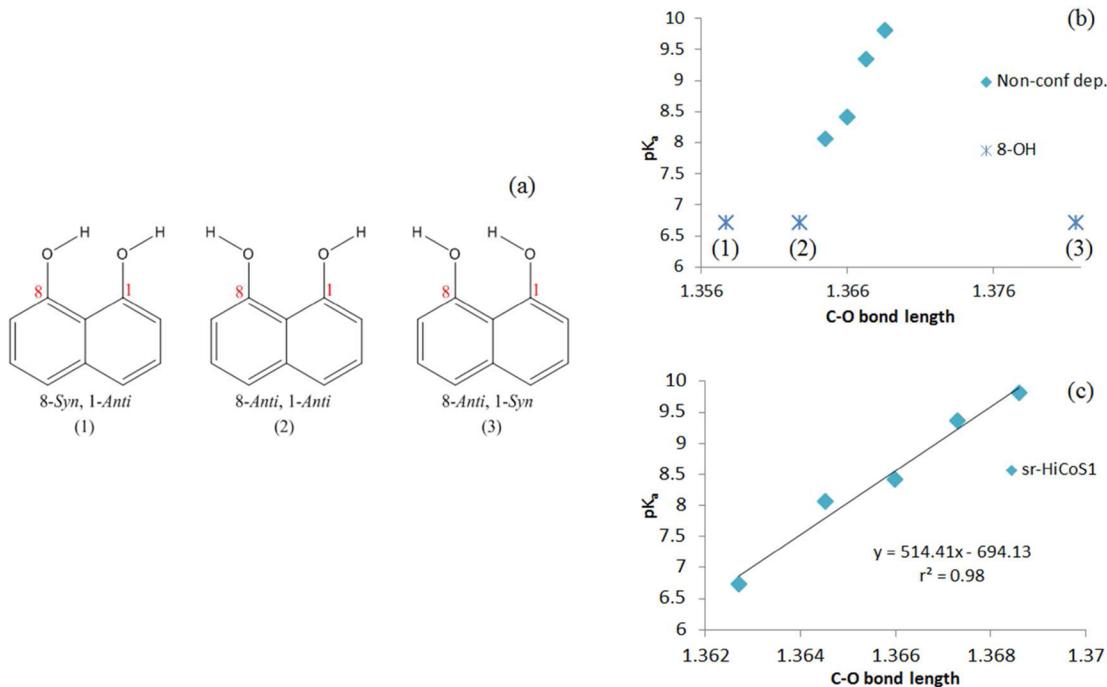


Figure 2. (a) Three of the four ($=2 \times 2$) possible conformers (excluding 8-syn,1-syn) of 1,8-dihydroxynaphthalene. (b) Active 1-OH C–O bond lengths corresponding to the three conformers. It should be noted that the predicted HiCoS selects conformer (2) as the correct conformation, and (c) shows its incorporation into the sr-HiCoS.

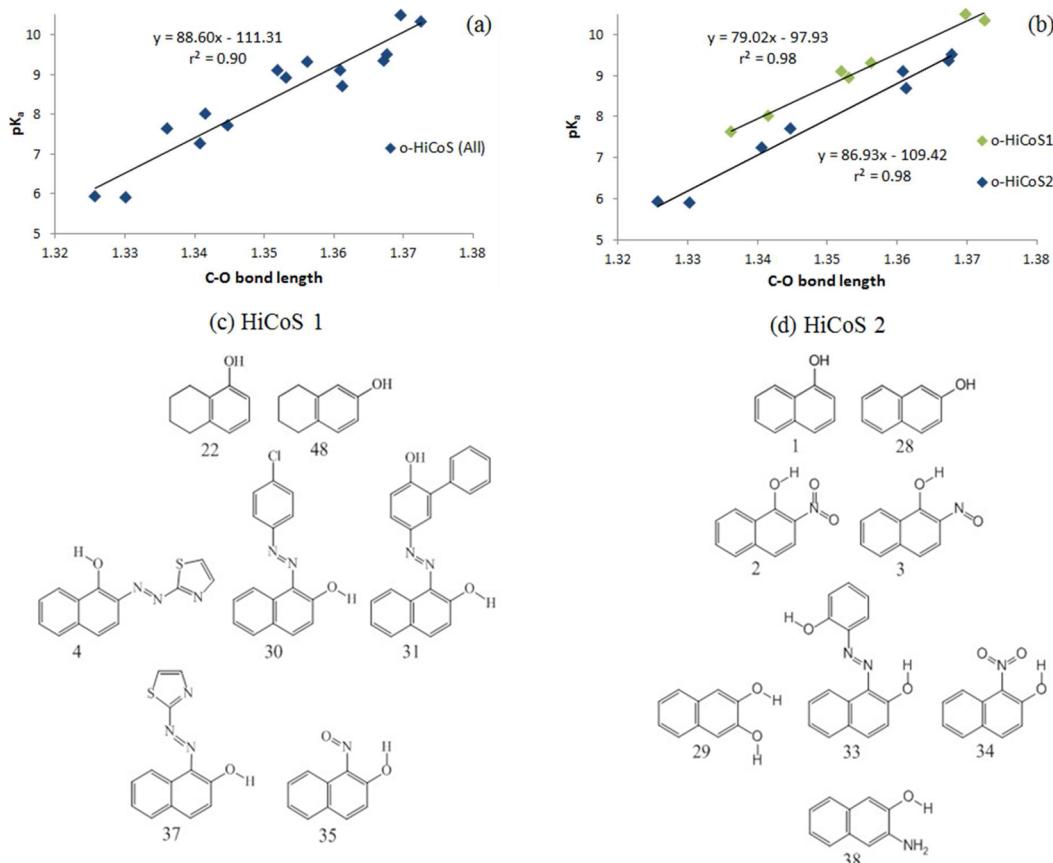


Figure 3. (a) Initial HiCoS for *o*-naphthols, shown as a correlation between the ab initio C–O bond length and the experimental pK_a . (b) Final and conformationally dependent HiCoSs, designated as o-HiCoS 1 and o-HiCoS 2. (c, d) Conformations preferred by the compounds in (c) o-HiCos 1 and (d) o-HiCoS 2.

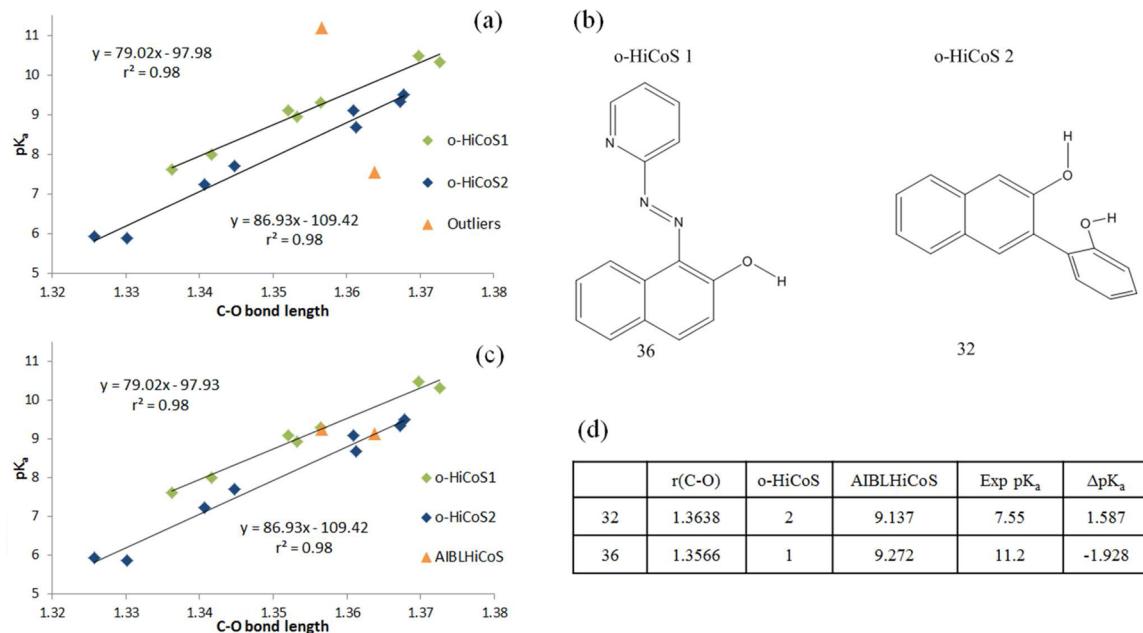


Figure 4. (a) The two *o*-HiCoSs with the two outliers to both data sets. (b) Structures of the outliers. (c) Corrected AIBLHiCoS $\text{p}K_a$ values and (d) numerical values of the outlier correction.

study was verified through experimental data and detailed a trend for *o*-phenols that rationally extends to the naphthol family studied. Naphthols are related to phenols through extension of conjugation via a second ring, which lowers the relative $\text{p}K_a$ values through resonance stabilization. The increased acidity observed for phenols with an intramolecular hydrogen bond should then be realized in the chemically similar naphthol family when hydrogen-bonding substituents are in close proximity.

3.2.2. Second Ring Case Study. There is a limited bank of experimental $\text{p}K_a$ values for substituents at positions in the *second ring* (sr) of 1-naphthol. **Table S4** lists the substituents of this new set, which consists of six data points including the 1-naphthol (compound 1) reference point. When all six data points are considered together, there are two obvious outliers, corresponding to the charged sulfur compounds 3,6-(SO_3^-)₂- and 5-(SO_3^-)-1-naphthol, which are expected to be outliers because of their anionic character (see the last paragraph in **section 2**). Using CPCM, again at the B3LYP/6-31G(d,p) level, allows these points to form part of the second ring HiCoS (sr-HiCoS). Reassignment of 3,6-(SO_3^-)₂-1-naphthol to the m+p-HiCoS improves both the new and old subsets, demonstrating that the effect of the substituent on the *primary* ring is dominant.

There is only one conformationally dependent species within the two families of naphthols, 1,8-dihydroxynaphthalene, a species with two functional groups in proximity, in which the conformation (*syn/anti*) could lead to the formation of an intramolecular hydrogen bond, drastically affecting the bond length and consequently the quality of the correlation with the experimental $\text{p}K_a$. Three from a total of four possible conformations were optimized and analyzed with respect to the predicted HiCoS (**Figure 2**). The fourth possibility is the conformation in which both groups are *syn* to one another, but this conformation has an imaginary frequency, understood to arise from a steric clash of hydrogen atoms. Because of the close proximity of the two alcohol groups, moving between

hydrogen-bonding and non-hydrogen-bonding conformers greatly perturbs the bond length of the active C–O bond. **Figure 2b** visualizes the effect of the conformation on the bond length, with a drastic change between conformers (1) and (3), in which the active alcohol (1-OH) behaves as a hydrogen-bond acceptor and donor, respectively. Conformer (2) in **Figure 2** (8-*anti*,1-*anti*) is the only one out of the three possible geometries that shows a clear correlation with the other data, demonstrating the importance of conformation when considering $\text{p}K_a$. Conformer (2) exhibits no intramolecular hydrogen bonding, and hence, $r(\text{C}-\text{O})$ is not affected by the rearrangement of electron density associated with this interaction. This allows the calculated bond length for (2) to fall within a set consisting of other second-ring-substituted compounds, which also exhibit no hydrogen-bonding interactions. The dramatic sensitivity to the conformation of the substituents amounts to self-selection. This fact not only demonstrates the importance of this parameter but also renders a small data set meaningful. Details of this HiCoS are given in **Table S5**.

3.3. *o*-Naphthols (*o*-HiCoS). The conformation adopted by all 15 compounds in the *ortho* HiCoS (*o*-HiCoS), listed in **Table S6**, has a drastic effect on the correlation with experimental $\text{p}K_a$ values. Through analysis of all of the conformations that are energetic minima it was possible to find the active bonds that gave a high correlation for this subset (**Figure 3**). Consideration of the lowest-energy conformers, all 15 of which are shown in **Figure 3a**, made it possible to establish two further subsets within the *o*-HiCoS, as shown in **Figure 3b**. The two HiCoSs are, with the sole exception of compound 35, split by whether the alcohol group is *syn* or *anti* to the second substituent. **Figure 3c,d** shows the geometries of the compounds in the two HiCoSs. The formation of these additional HiCoSs drastically improves the correlations, as demonstrated in **Figure 3b**. Further details on the two *o*-HiCoSs can be found in **Tables S7 and S8**.

“Self-selection” describes how chemical space is further partitioned according to the preferred gas-phase conformation

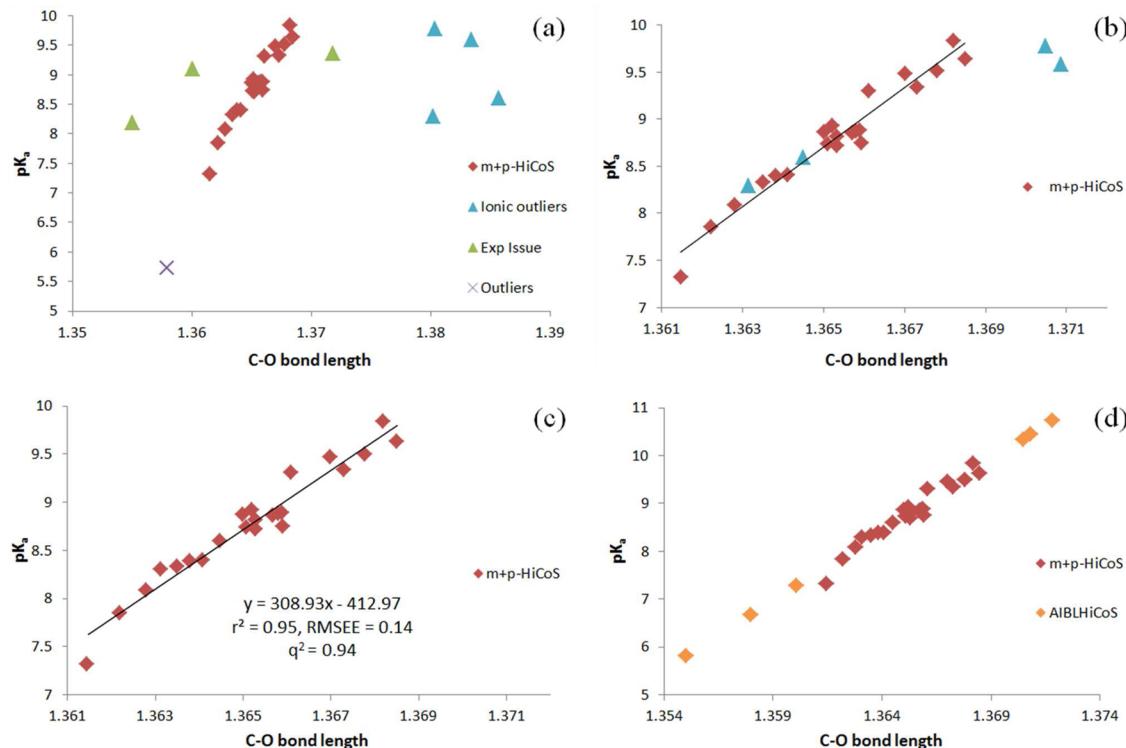


Figure 5. (a) Initial set of *m*- and *p*-naphthols, shown as a correlation between the ab initio C–O bond length and the experimental pK_a , with three types of outliers highlighted. (b) Correction of the four ionic outliers through implementation of CPCM, allowing two of the blue triangles in (a) to join the rest of the set. The three points assigned to the experimental issues section plus the one remaining outlier (4-NO_2) shown in (a) have been removed in (b). Additionally, the two data points [$3\text{-(COO}^-)$ and $4\text{-(COO}^-)$] still remained outliers for reasons discussed in the main text and were removed from the correlation. (c) *m+p*-HiCoS and the corresponding AIBLHiCoS equation. (d) *m+p*-HiCoS extended by AIBLHiCoS-corrected remaining outliers.

of the electronically congeneric compounds of each set. However, in the case of the *o*-naphthols we are faced with an exception in the form of 1-NO-2-naphthol (35 in Figure 3). The defining feature of the compounds of *o*-HiCoS 1 is that they all have a gas-phase global minimum geometry exhibiting an *anti* conformation (disregarding reference compounds 22 and 48). However, for compound 35, the calculated bond length means that it falls into neither subset when classified by its conformation. The most stable structure in the gas phase exhibits intramolecular hydrogen bonding, as one might expect in the absence of any explicit solvation, between the hydroxyl proton and the oxygen of the nitroso group (akin to compound 34) and falls into neither set. Instead, 35 appears as an outlier with respect to the two lines of HiCoSs 1 and 2, with a much shorter $r(\text{C}-\text{O})$ than the rest of the set. In the *anti* conformation, the system is stabilized by hydrogen bonding, this time with the hydroxyl acting as a hydrogen-bond donor to the nitrogen of the nitroso group, with the $r(\text{C}-\text{O})$ allowing it to fit within HiCoS 1. The highest-energy *anti* structure exhibits no hydrogen bonding and fits within HiCoS 2, where all of the other members are classified as *syn* conformers. With all of the possible conformations being considered and given the proven power of our method for structurally similar compounds within the set, especially considering compound 3 (2-NO-1-naphthol), it seems that there may be some degree of error in the experimental pK_a .

The seven conformers that make up *o*-HiCoS 1 have proportionally higher pK_a values for a similar range of bond lengths than the eight compounds of *o*-HiCoS 2 and thus are inherently more acidic. The reference points used to shape *o*-

HiCoS 1, namely, compounds 22 and 48, the two naphthols in which the second ring is saturated, are phenolic in nature. The carbons of the second ring for these two compounds are thus sp^3 hybridised, so the extent of conjugation is restricted to that of a phenol. However, the reference points used for the *o*-HiCoS 2, compounds 1 and 28, are naphtholic in nature. Upon consideration of the preferential conformation for the compounds in *o*-HiCoS 2, it is observed that each alcohol group is *syn* to the *ortho* substituent, corresponding to the conformation that can form internal hydrogen bonds. In most *o*-phenols, the acidity of the O–H group is increased upon formation of intramolecular hydrogen bonds³⁹ as a result of rearrangement of electron density in the molecule. This trend is indeed observed for *o*-naphthols in conformers that can form intramolecular hydrogen bonds: through the formation of internal hydrogen bonds, the compounds in *o*-HiCoS 2 exhibit heightened acidity, reflected in the relatively lower pK_a values than for *o*-HiCoS 1.

For both *o*-HiCoSs, two outliers were found (compounds 32 and 36; Table S6). Figure 4a shows the outliers with respect to the two established *o*-HiCoSs. The two conformers (out of the four possibilities discussed in connection with Figure 2) that gave the closest fits to the observed correlations, shown in Figure 4b, exhibit conformational traits that allow them to be confidently assigned an *o*-HiCoS. It is pleasing to note that the principle behind the self-selection of conformers again applies here. The experimental pK_a values reported for these two compounds were measured at a nonstandard temperature with respect to the other compounds in the HiCoSs, indicating that

erroneous results rather than selection of the correct conformation is to blame for the poor correlation.

The temperatures at which these two experimental pK_a measurements were made were 20 and 30–36 °C for compounds 32 and 36, respectively. The temperature differences of −5 and +5–11 °C with respect to the standard temperature of 25 °C for the o-HiCoSs always have a drastic effect on the recorded pK_a values. Within the current o-HiCoSs, the temperature changes are understood to have led to an almost uniform rising and lowering of the relative pK_a values with respect to a lowering and rising of temperature. The conformer that gave the best results for these compounds allowed the assignment of compound 36 (the one at the top of Figure 4a) to o-HiCoS 1 and compound 32 to o-HiCoS 2. In particular, the equation in Table S7 (for o-HiCoS 1) and the one in Table S8 (for o-HiCoS 2) predicted the correct pK_a values for these compounds. The predictions raised the pK_a of compound 32 by 1.59 units and lowered that of compound 36 by 1.93 units. It is possible to estimate the pK_a of a compound at standard temperature by implementation of the van't Hoff equation, which describes the response of the equilibrium constant (in this case for aqueous proton dissociation) to a change in temperature. In order to approximate the change in the pK_a from nonstandard to standard temperature, the enthalpy change for the dissociation of the proton under aqueous conditions must be known. Using the procedure outlined in the last section of the SI, for compound 32 we obtained a predicted pK_a of 7.85, an increase of 0.30 pK_a units from the original value of 7.55 at 30–36 °C. For the calculation we assigned a temperature of 33 °C (at the midpoint between 30 and 36 °C). This increase in pK_a with a decrease in temperature of 8 °C is much more subtle than that obtained using the equation of our line (0.30 using van't Hoff and 1.59 using the o-HiCoS 2 line). For compound 36, the increase in temperature by 5 °C was found to cause a decrease in pK_a from 11.20 to 11.13. This again suggests a much more subtle difference in pK_a than obtained using the equation of the line for o-HiCoS 2 (−0.07 for van't Hoff calculations versus −1.93 using our method). It could be suggested that the method for calculating the enthalpy change for proton dissociation in aqueous solution would suffer from the inadequacies of the CPCM model in reproducing the stability afforded by the presence of solvent molecules. As there is experimental data available for the standard conditions, we are unable to make comparisons to determine which method is more accurate.

3.4. Subsets. *3.4.1. m+p-HiCoS.* The *meta* and *para* substituents were first considered as two subsets and then later combined to give one larger subset with an improved spread of data. The correlation observed for the initial subset (Table S9) was poor because of obvious outliers, some of which could be removed with chemical intuition.

The effect of electron-withdrawing and -donating groups on the bond dissociation enthalpy of phenolic acidic protons has been explored previously in the literature,³⁶ with an opposite effect seen for the two classes of substituents. Electron-donating groups, such as OMe, weaken the O–H bond in phenols, whereas electron-withdrawing groups, such as NO₂, cause an increase in the bond dissociation energy (BDE) of the O–H bond. This is in contrast to the negligible effect of electron-donating substituents at the *meta* position of phenol, whereas electron-withdrawing groups at the *meta* position show the same trend, giving increased BDE(O–H) values.

For all of the optimizations in the gas phase, the compounds with the ionic substituents $-SO_3^-$ and $-CO_2^-$, shown as blue triangles in Figure 5a, turned out to be outliers for all of the basis sets used. Utilization of an implicit-solvent model, CPCM, at the modest level of theory B3LYP/6-31G(d,p) gave a much better representation of the ionic systems, allowing consideration of all of the compounds in their predicted HiCoS (Figure 5b). The sulfonate-containing naphthols fit with the correlation of their respective data sets with much greater accuracy than the carboxylate-containing naphthols. The acidities of sulfonates are sufficiently low to regard this group as essentially always unprotonated,⁴⁰ while the naphtholic alcohol group is protonated. In other words, the microscopic dissociation values of the sulfonate group and the alcohol group are distinct enough that the differences between these (microscopic) values, unique to each functional group, and the macroscopic dissociation values become negligible. Our methodology uses a descriptor (i.e., a bond length) that is unique to a microstate structure and therefore is accurate only when predicting the corresponding microscopic dissociation constant. In the case of the sulfonate-containing naphthols, multi-deprotonation can be regarded as a truly stepwise process and the microstate-optimized pK_a matches the macroscopic dissociation constant reported experimentally. This difference in microscopic dissociation constants of functional groups is not as distinct when looking at the values for a carboxylic acid and an alcohol. Therefore, the macroscopic dissociation constant reported will be affected in a non-negligible way by a mixing of microstate structures in which a proton is exchanged from the alcohol to the carboxylate group. As our methodology does not yet account for such simultaneous and multiple contributions of microscopic dissociation constants to the macroscopic constant, it is to be expected that these results will appear as outliers regardless of the level of theory used in optimization. As the AIBLHiCoS model depends on the active bond length of a microstate and a macroscopic dissociation constant, the outlier becomes easier to understand. The bond length provided by the microstate allows prediction of the microscopic pK_a value for each of the carboxylic acid outliers (Figure 5d).

A second class of outliers (compounds 14, 15, and 17 in Table S9) are shown in Figure 5a as green triangles. These points represent compounds that are subject to an experimental issue associated with their pK_a values. One of these compounds, 4-azopyridine-1-naphthol (15), was measured at the non-standard temperature of 30–36 °C, which is 5–11 °C higher than the standard conditions reported for the other measurements. According to Le Chatelier's principle, which states that an increase in temperature corresponds to a shift in the endothermic direction of the equilibrium, an increase in the pK_a with a decrease in temperature is to be expected. Unfortunately, on this occasion our method fails to give a realistic pK_a at 25 °C for this compound, as use of the calculated $r(C-O)$ in the equation of the m+p-HiCoS gives us a pK_a value of 7.31. This is several units lower than expected for a decrease in temperature by 8 °C. The van't Hoff calculations give an estimated pK_a at 25 °C of 9.37, or 0.27 units higher than the original value. Experimental data at standard temperature is not available for comparison.

The second outlier corresponds to 4-NO-1-naphthol (14), but here there a history of doubt as to the exact identity of the nitrosonaphthol measured by Trubbsbach.⁴¹

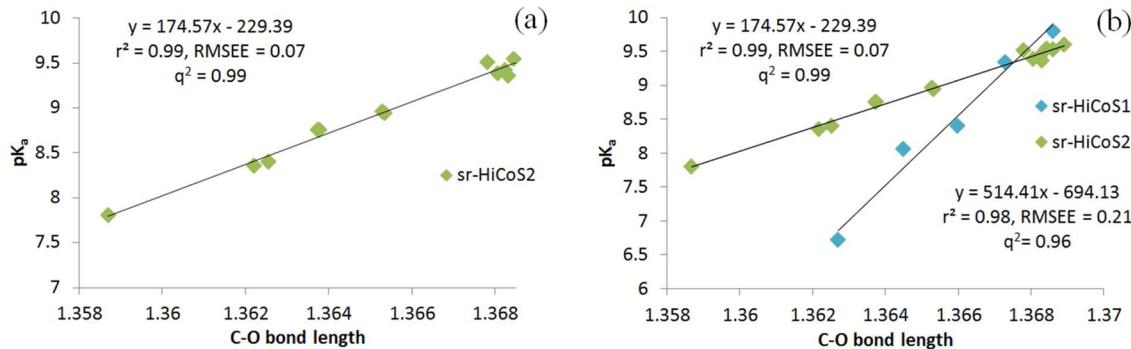


Figure 6. (a) HiCoS for 14 second ring (sr) 2-naphthols, shown as a correlation between the ab initio C–O bond length and experimental pK_a . (b) HiCoSs for the 1- and 2-naphthol second ring substituents (sr-HiCoS 1 and sr-HiCoS 2, respectively).

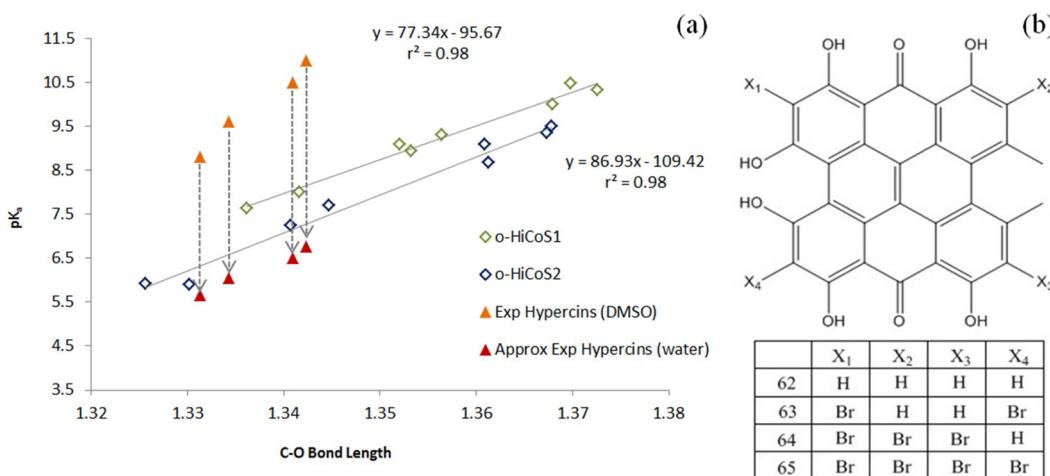


Figure 7. (a) The o-HiCoSs in water and the correlation for the hypericin-based biomolecules in DMSO, shown as a correlation of the ab initio C–O bond length and experimental pK_a . (b) Skeletal structure of the hypericin-based class of compounds.

The final outlier, 4-OH-1-naphthol (17), was measured at 26.4 °C, so we can expect the pK_a value to be marginally lower than if it were measured at standard conditions. Using the equation of our line for the m+p-HiCos, we obtain a predicted pK_a of 10.73, which is in agreement when (tentatively) compared to the value of 10.26 ± 0.40 predicted using the ACD Laboratories methodology. The van't Hoff calculation yields a pK_a of 9.42 for the decrease in temperature of 1.4 °C, which is 0.05 units higher than the original value. Because of the high structural similarity of 4-OH-1-naphthol to the rest of the set and indeed the fact that the calculated bond length for 3-OH-1-naphthol allows it to fall nicely on the line, it is perhaps sensible to consider the original experimental pK_a value as erroneous.

Removal of the three compounds with an experimental issue and the two charged carboxylate-containing naphthols would give a much improved correlation and a promising HiCoS. However, there still remains one outlier, 4-NO₂-1-naphthol (compound 13 in Table S9, represented by the cross in Figure 5a), removal of which (as shown in Figure 5c) gives a subset with a high correlation from which outliers can be corrected.

Removal of all outliers gave the m+p-HiCoS, shown in Figure 5c and Table S10, with an impressive correlation for the HiCoS. For all of the remaining outliers, the AIBLHiCoS model provided an equation of correlation that was used to predict an accurate pK_a . The AIBLHiCoS-corrected pK_a values are shown in Figure 5d.

3.4.2. sr-HiCoSs. As demonstrated above, there is a clear and well-defined sr-HiCoS for the 1-naphthols. There is a larger set of 14 experimental pK_a values for the 2-naphthols,^{42–45} given in Table S12. Because of the position of the alcohol group, there are no conformational effects that must be taken into account, and the enhanced distance between the two substituents is seen to produce a much shallower slope and hence higher pK_a values than those of the 1-naphthol sr-HiCoS (Figure 6b). The two HiCoSs for the second rings of the 1- and 2-naphthols cannot be considered together, as the proximity of the two substituents in the 1-naphthols is shown to have a much greater effect, with the corresponding enhanced acidity and steeper linear dependence on the bond length.

3.5. External Test Set. 3.5.1. Biologically Relevant Molecules. AIBLHiCoS provides a pK_a prediction method that can accurately predict dissociation constants of larger biomolecules following the establishment of trends for smaller fragments found in the drug compound. Our work so far has demonstrated the clear link between the phenol and naphthol families, a link that can be exploited to explore the larger aromatic backbones of many active drug compounds. Our findings below demonstrate not only the novel ability of our methodology to identify “active fragments” of larger biomolecules but also how this enables us to predict pK_a values more accurately than packages currently employed.

Benzoid aromatic fragments are a common motif in dyes and medicinal compounds. The conjugated-π-bond chromo-

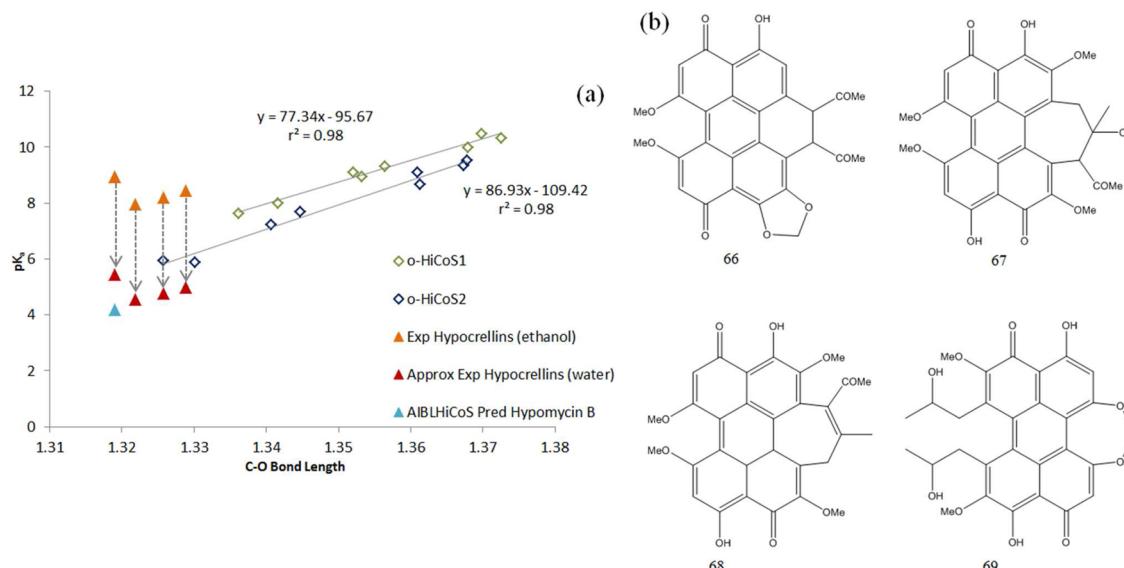


Figure 8. (a) The o-HiCoSs and the correlation for the hypocrellin-based biomolecules, shown as a correlation between the ab initio C–O bond length and experimental pK_a in water, the literature experimental pK_a values measured in ethanol (orange triangles) and the new experimental pK_a values (red triangles) approximated using the linear relationship between the experimental pK_a values in ethanol and water, and finally the AIBLHiCoS-predicted pK_a for hypomycin B (66) (blue triangle). (b) Skeletal structures of the hypocrellin-based class of compounds.

phoric fragment of these compounds produces vivid colors in dyes and is the photosensitizer used in photodynamic therapy.^{46,47} Some ambiguity exists in extended conjugated structures as to whether the active (benzenoid) fragment corresponds to the phenol, naphthol, or anthrol family. However, there is a clear link between these three classes that will indeed allow clear identification of the correct family and HiCos.

Experimental values of the biologically relevant molecules used to test the naphthol HiCos were measured in a variety of solvents. The influence of solvation has often been obscured by the dominance of water as a solvent in quantitative studies of acid–base properties. However, there is plenty of evidence showing that there is often a marked difference of behavior in nonaqueous solvents compared with water. Exploration of the relationships between pK_a scales in water and other aqueous solvents shows rational and quantitatively predictable trends.⁴⁸ Good linear correlations between pK_a values give equations specific to chemical families, and from these equations follow accurate predictions of pK_a values in nonaqueous solvents from the pK_a value in water.

We will now demonstrate the practical use of AIBLHiCoS on four classes of biomolecules: hypericins, hypocrellins, and chromomycinones, all predicted via the o-HiCoS, and finally estrogens, predicted from the sr-HiCoS and phenol m+p-HiCos.

3.5.2. o-HiCos. To be able to evaluate the relationship between the HiCos and the biological molecules tested, relevant previously established linear relationships⁴⁸ were used to transform the pK_a values of the biomolecules from the experimental solvent systems to water. Table S13 gives the equations used for both transformations and the resultant approximate pK_a values. The pK_a values for the first class of four biological molecules, consisting of hypericin and three brominated hypericin derivatives, were originally measured in dimethyl sulfoxide (DMSO), a solvent that drastically raises the pK_a values of phenols. Figure 7a shows a clear trend when the experimental DMSO-scale pK_a values of the hypericin

compounds are plotted alongside the water-scale pK_a values of the o-HiCos. The hypericin compounds in DMSO form a subset that has a range of pK_a values much higher than that of the naphthols in aqueous solution. This observation can be explained by the inadequacy of the less polar DMSO molecules at forming a solvation shell around the anion compared to water. The hypericin structure is more extensively conjugated than that of the phenol and naphthol families, a structural feature that has been demonstrated to heighten the acidity in going from the phenol family to the naphthol family. This leads to an intuitive prediction that for the extensively conjugated biomolecules there will again be a lowering of the pK_a values relative to those of the naphthol and phenol families due to greater stabilization of the conjugate base. Upon conversion of all of the values to the water or DMSO scale, the hypericin class's correlation falls in the predicted place. Figure 7a shows the experimental pK_a values for the hypericin class converted to the water scale, confirming the above prediction.

Figure 8 details a second class of four biomolecules whose pK_a values were measured in a nonaqueous solvent system. This class, consisting of the compounds hypomycin B (compound 66), cercosporin (69) and hypocrellin A (67) and B (68), which share a perylene structural commonality, are shown in Figure 8b. The hypocrellin-based structures showed a high correlation at relatively higher pK_a values than predicted. Measurements of pK_a for these structures were done in ethanol, which is known to raise pK_a values with respect to the aqueous pK_a scale. Upon conversion of these experimental values to the water scale, the expected correlation was seen. There was a single outlier, hypomycin B (66), which had an experimental pK_a value far greater than that suggested by the bond length associated with the structure. The experimental value seems far greater than expected on the basis of the high level of structural similarity with the other compounds in this group. There is a difference of ~1 pK_a unit between the experimental values for cercosporin and hypomycin B, which is greater than the difference between the reference naphthols and a neighbor alcohol group. On the basis of the high correlation seen for the

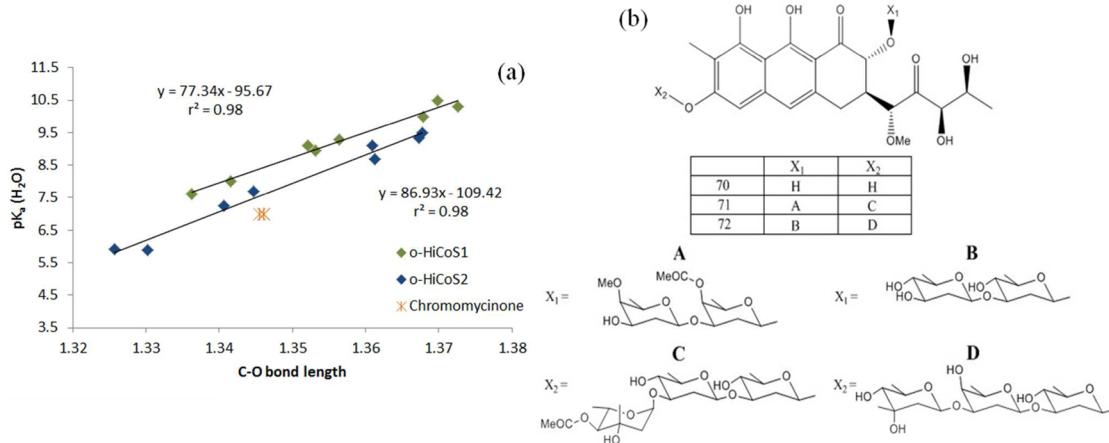


Figure 9. (a) The o-HiCoSs and the correlation for the chromomycinone-based biomolecules, shown as a correlation between the ab initio C–O bond length and experimental pK_a in water. (b) Skeletal structure of the chromomycinone-based class of compounds.

other species in this class and the systematic dependence on the bond length observed, it is prudent to regard this experimental measurement as erroneous. Through the AIBLHiCoS methodology it is possible to suggest a value of 4.18, which lies 1.28 pK_a units lower than the solvent-corrected value of 5.44.

A third class of common active skeletons, including chromomycinone, plicamycin, and chromomycin A3, which differ only in the saccharide side chain, is shown in Figure 9. All of these compounds have the same experimental pK_a values, demonstrating that the active fragment of the molecule, in these cases at least, can be understood to be restricted to the conjugated chromomycinone skeleton.

Figure 10 shows the correlation of the three classes of biomolecules considered separately above. Through correction

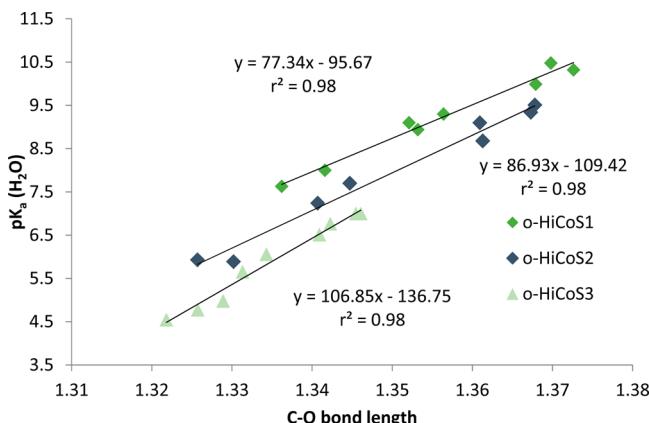


Figure 10. o-HiCoSs for naphthol and the biomolecules, shown as correlations between the ab initio C–O bond length and experimental pK_a in water.

of the experimental pK_a scales for the different classes of biomolecules it is possible to obtain a data set with a good correlation that can be understood in the context of the previously discussed o-HiCoSs. The biomolecules form an o-HiCoS of their own, o-HiCoS 3, given in Table S14. The pK_a values of these larger aromatic compounds forming o-HiCoS 3 are more acidic than those of the naphthol o-HiCoSs. The heightened acidity was predicted through consideration of the increased conjugation and electronegative neighboring atoms in this set.

Table 1 provides the predicted pK_a values for the compounds used to form o-HiCoS 3. AIBLHiCoS predictions were made using the equation given in Table S14. For each of the prediction methodologies, the variation between the experimental and predicted pK_a values is given. The AIBLHiCoS predictions are consistently more accurate than either of the alternative predictions, with the greatest error associated with compound 68 (0.26 units). For the non-AIBLHiCoS methodologies, the errors range from small to glaring. It is unclear whether these predictions arise from overestimation of the effect upon addition of substituents around the active fragment or from lack of compensation for nonaqueous solvent systems in the design of the package. Through consideration of both of the above factors we present a prediction method that can negotiate these complexities.

3.5.3. sr-HiCoSs. The biomolecules above all fit into the o-HiCoS with good correlation. There were compounds that showed a high correlation with other HiCoSs. A number of experimental pK_a values were available for estrogen steroids,⁴⁹ which have high structural similarity with m+p-phenols or sr-naphthols; these are given in Table S15. It was predicted that the two compounds with high structural similarity to the sr-naphthol compounds, equilenin and 17 β -dihydroequilenin, would fit the 2-naphthol sr-HiCoS. The other seven estrogen derivatives had much greater structural similarity to the m+p-phenol-HiCoS. The degree of structural similarity is based on the extent of conjugation with respect to the active C–O bond (near the top right of each molecule), which dictates whether the molecule belongs in the m+p-HiCoS of the phenol family or the sr-HiCoS of the naphthol family. The phenol m+p-HiCoS is discussed in greater detail below in the context of both the naphthol m+p-HiCoS and the “phenolic” estrogen biomolecules.

Figure 11 shows the correlation of the naphtholic estrogens with the sr-HiCoSs. As the hydroxyl group is at the 2-position of the naphthol skeleton, it was predicted that these two estrogen compounds would have a greater correlation with the sr-HiCoS for the 2-naphthols. The graph indicates that the dependence relationship between the bond length and the experimental pK_a shows similarity to that seen in the predicted HiCoS. The two values sit at slightly higher values of pK_a than predicted using AIBLHiCoS, by 0.35 units, a discrepancy that can be explained by a variety of subtle changes to the experimental setup that result in minor fluctuations in pK_a .

Table 1. Experimental Values of pK_a in Aqueous Solution for the Drug Compounds That Form o-HiCoS 3, Compared with the Predictions of AIBLHiCoS and Two Prediction Packages

| compd no. | biomolecule | exptl pK_a | AIBLHiCoS | ΔpK_a | ACE and JChem Acidity Calculator | ΔpK_a | ChemAxon | ΔpK_a |
|-----------|---------------------|--------------|-----------|---------------|----------------------------------|---------------|----------|---------------|
| 62 | hypericin | 6.77 | 6.673 | -0.097 | 8.9 | 2.13 | 6.91 | 0.14 |
| 63 | dibromohypericin | 6.52 | 6.527 | 0.007 | 7.6 | 1.08 | 4.96 | -1.56 |
| 64 | tribromohypericin | 6.06 | 5.820 | -0.240 | 7.6 | 1.54 | 4.82 | -1.24 |
| 65 | tetrabromohypericin | 5.66 | 5.503 | -0.157 | 7.2 | 1.54 | 4.24 | -1.42 |
| 67 | hypocrellin A | 4.78 | 4.904 | 0.124 | 9.1 | 4.32 | 7.03 | 2.25 |
| 68 | hypocrellin B | 4.98 | 5.242 | 0.262 | 8.7 | 3.72 | 6.62 | 1.64 |
| 69 | cercosporin | 4.55 | 4.485 | -0.065 | 9.2 | 4.65 | 7.25 | 2.70 |
| 70 | chromomycinone | 7 | 7.012 | 0.012 | 9.1 | 2.10 | 4.74 | -2.26 |
| 71 | chromomycin A3 | 7 | 7.079 | 0.079 | 7.9 | 0.90 | 4.54 | -2.46 |
| 72 | plicamycin | 7 | 7.079 | 0.079 | 7.9 | 0.90 | 4.54 | -2.46 |

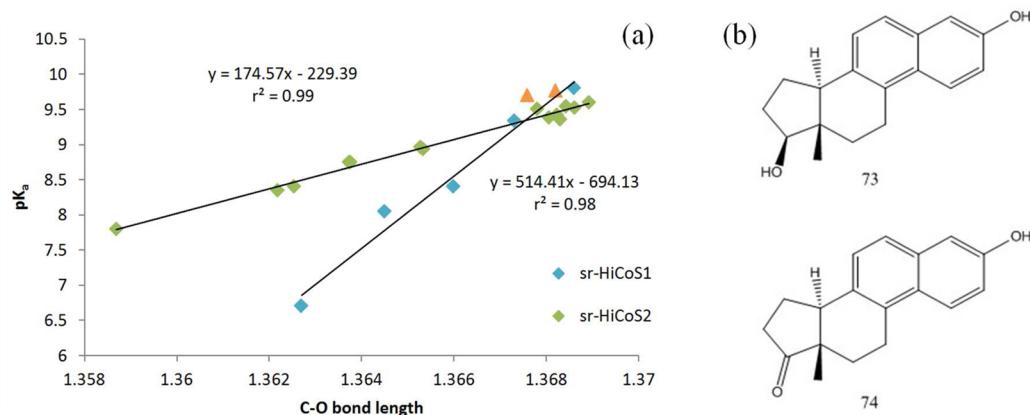


Figure 11. (a) The sr-HiCoSs and the active bond lengths for the naphtholic estrogen compounds (orange triangles), shown as a correlation between the ab initio C–O bond length and experimental pK_a . (b) Skeletal structures of the naphtholic estrogen compounds.

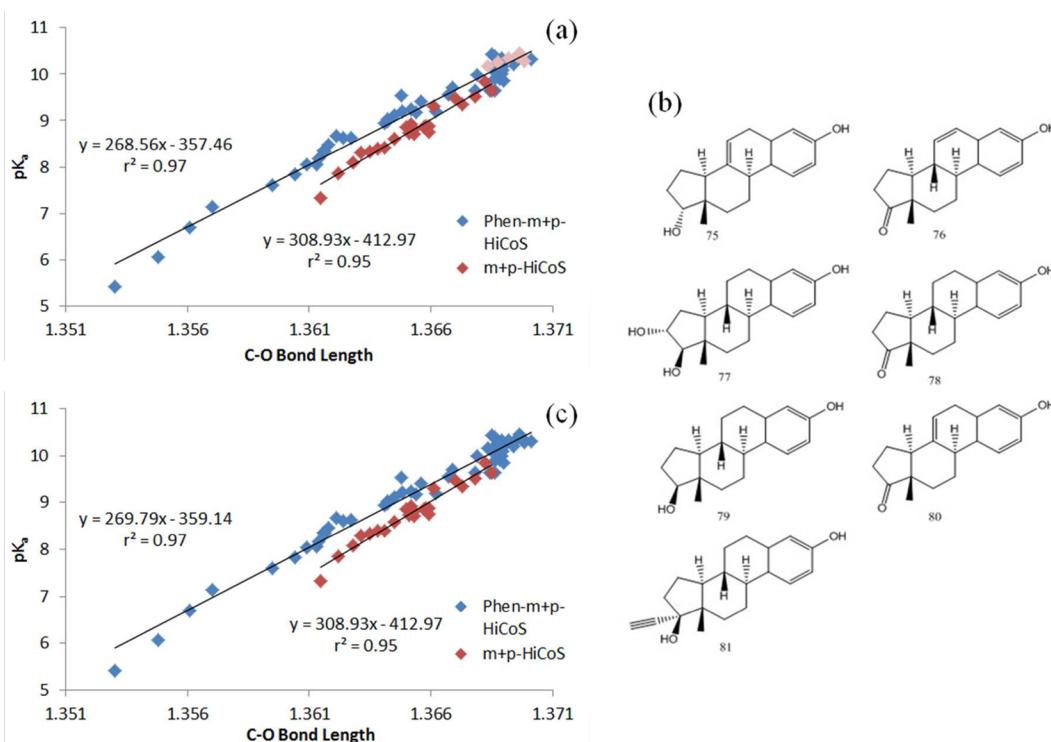


Figure 12. (a) The phenol and naphthol m+p-HiCoSs and the correlation for the phenolic estrogen compounds, shown as a correlation between the ab initio C–O bond length and experimental pK_a . (b) Skeletal structures of the phenolic estrogens. (c) Incorporation of the biomolecules into the phenol m+p-HiCoS.

measurements. Regardless of the absolute experimental pK_a value reported, the relationship between the two compounds and the predicted HiCoS shows a strong linear correlation.

3.5.4. Phenolic HiCoS Biomolecules. As mentioned in the previous section, seven estrogen derivatives were predicted on the basis of structural commonality to fit the phenol m+p-HiCoS with a strong correlation (Figure 12). The phenolic estrogens fit the predicted HiCoS with high accuracy. The strong results shown regarding the predictability of HiCoSs across families of chemical compounds give great weight to the methodology. Through the comparative work done with the phenol family studied previously and the naphthol family of this study, it has been possible to establish common HiCoSs based on chemical space and to understand which fragment of the molecule is active in a larger biomolecule with regard to the extent of conjugation. The relationships shown in this body of work are chemically intuitive. It is understood and predicted that with more conjugation there is greater stabilization of the conjugate base, which lowers the pK_a value. It is also natural that with a closer chemical proximity between the active functional group and the secondary substituent, a greater influence on the pK_a will emerge.

CONCLUSION

We have presented a simple pK_a prediction method called AIBLHiCoS based on a single parameter, namely, an “active bond” in families of compounds with chemical commonality. The corresponding model is tightly linked to experiment and possesses the ability to correct erroneous results and determine the correct microscopic dissociation constant. Case studies on large aromatic biomolecules demonstrate the powerful relationships found for commonly and chemically intuitive High Correlation Subsets (HiCoSs) as seen in phenols and naphthols. These relationships appear not only within chemically similar families but also between the chemical families themselves. Through an understanding of the relationship between aromatic compounds with one and two rings, AIBLHiCoS can be generalized to accurately predict trends through extension of the delocalized system.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jcim.5b00580](https://doi.org/10.1021/acs.jcim.5b00580).

Tables S1–S17, Figures S1 and S2, and treatment of the temperature dependence with the van't Hoff equation ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

*E-mail: paul.popelier@manchester.ac.uk. Phone: +44 161 3064511.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.A. and B.A.C. thank the BBSRC and Syngenta Ltd. for funding their Ph.D. studentships.

REFERENCES

- (1) Ugur, I.; Marion, A.; Parant, S.; Jensen, J. H.; Monard, G. *J. Chem. Inf. Model.* **2014**, *54*, 2200–2213.
- (2) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 704–721.
- (3) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Struct., Funct., Genet.* **2008**, *73*, 765–783.
- (4) Krieger, E.; Nielsen, J. E.; Spronk, C. A. E. M.; Vriend, G. *J. Mol. Graphics Modell.* **2006**, *25*, 481–486.
- (5) Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89*, 141–157.
- (6) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 738–752.
- (7) Meng, Y. L.; Dashti, D. S.; Roitberg, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 2721–2727.
- (8) Shi, C. Y.; Wallace, J. A.; Shen, J. K. *Biophys. J.* **2012**, *102*, 1590–1597.
- (9) Wallace, J. A.; Shen, J. K. *Methods Enzymol.* **2009**, *466*, 455–75.
- (10) Liton, M. A. K.; Ali, M. I.; Hossain, M. T. *Comput. Theor. Chem.* **2012**, *999*, 1–6.
- (11) Angel, L. A.; Ervin, K. M. *J. Phys. Chem. A* **2006**, *110*, 10392–10403.
- (12) da Silva, G.; Kennedy, E. M.; Dlugogorski, B. Z. *J. Phys. Chem. A* **2006**, *110*, 11371–11376.
- (13) Dissanayake, D. P.; Senthilnithy, R. *J. Mol. Struct.: THEOCHEM* **2009**, *910*, 93–98.
- (14) Ghalami-Choobar, B.; Dezhampahan, H.; Nikparsa, P.; Ghiamishomami, A. *Int. J. Quantum Chem.* **2012**, *112*, 2275–2280.
- (15) Zielinski, F.; Tognetti, V.; Joubert, L. *J. Mol. Model.* **2013**, *19*, 4049–58.
- (16) Cheng, J.; Sulpizi, M.; Sprik, M. *J. Chem. Phys.* **2009**, *131*, 154504.
- (17) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
- (18) Antosiewicz, J.; Mccammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.
- (19) Casasnovas, R.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Munoz, F. *Int. J. Quantum Chem.* **2014**, *114*, 1350–1363.
- (20) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; Garcia-Moreno, B. *Biophys. J.* **2002**, *82*, 3289–3304.
- (21) Fraczkiewicz, R.; Lobell, M.; Goller, A. H.; Krenz, U.; Schoenneis, R.; Clark, R. D.; Hillisch, A. *J. Chem. Inf. Model.* **2015**, *55*, 389–397.
- (22) Harding, A. P.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11283–11293.
- (23) Harding, A. P.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11264–11282.
- (24) Alkorta, I.; Griffiths, M. Z.; Popelier, P. L. A. *J. Phys. Org. Chem.* **2013**, *26*, 791–796.
- (25) Griffiths, M. Z.; Alkorta, I.; Popelier, P. L. A. *Mol. Inf.* **2013**, *32*, 363–376.
- (26) Alkorta, I.; Popelier, P. L. A. *ChemPhysChem* **2015**, *16*, 465–469.
- (27) Bryson, A.; Matthews, R. W. *Aust. J. Chem.* **1963**, *16*, 401–410.
- (28) Tolbert, L. M.; Solntsev, K. M. *Acc. Chem. Res.* **2002**, *35*, 19–27.
- (29) Agmon, N.; Rettig, W.; Groth, C. *J. Am. Chem. Soc.* **2002**, *124*, 1089–1096.
- (30) Houari, Y.; Jacquemin, D.; Laurent, A. D. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11875–11882.
- (31) Premont-Schwarz, M.; Barak, T.; Pines, D.; Nibbering, E. T. J.; Pines, E. *J. Phys. Chem. B* **2013**, *117*, 4594–4603.
- (32) Kowalik, J.; Van der Veer, D.; Clower, C.; Tolbert, L. M. *Chem. Commun.* **1999**, 2007–2008.
- (33) Cruciani, G.; Milletti, F.; Storchi, L.; Sforza, G.; Goracci, L. *Chem. Biodiversity* **2009**, *6*, 1812–1821.
- (34) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Phys.* **2008**, *128*, 144108.
- (35) Moran, D.; Simmonett, A. C.; Leach, F. E.; Allen, W. D.; Schleyer, P. V.; Schaefer, H. F. *J. Am. Chem. Soc.* **2006**, *128*, 9342–9343.
- (36) Chandra, A. K.; Uchimaru, T. *Int. J. Mol. Sci.* **2002**, *3*, 407–422.
- (37) Serjeant, E. P.; Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution*; Pergamon Press: Oxford, U.K., 1979.

- (38) Huque, F. T. T.; Platts, J. A. *Org. Biomol. Chem.* **2003**, *1*, 1419–1424.
- (39) Kanamori, D.; Furukawa, A.; Okamura, T.; Yamamoto, H.; Ueyama, N. *Org. Biomol. Chem.* **2005**, *3*, 1453–1459.
- (40) Henson, R. M. C.; Wyatt, P. A. H. *J. Chem. Soc., Faraday Trans. 2* **1975**, *71*, 669–681.
- (41) Creamer, L. K.; Packer, J.; Vaughan, J.; Fischer, A.; Richards, R. B.; Mann, B. R. *J. Org. Chem.* **1961**, *26*, 3148–3155.
- (42) Tolbert, L. M.; Haubrich, J. E. *J. Am. Chem. Soc.* **1994**, *116*, 10593–10600.
- (43) Tolbert, L. M.; Haubrich, J. E. *J. Am. Chem. Soc.* **1990**, *112*, 8163–8165.
- (44) Huppert, D.; Tolbert, L. M.; Linares Samaniego, S. *J. Phys. Chem. A* **1997**, *101*, 4602–4605.
- (45) Snyckers, F.; Zollinger, H. *Helv. Chim. Acta* **1970**, *53*, 1294–1305.
- (46) Dolmans, D. E. J. G. J.; Fukumura, D.; Jain, R. K. *Nat. Rev. Cancer* **2003**, *3*, 380–387.
- (47) Spring, B. Q.; Rizvi, I.; Xu, N.; Hasan, T. *Photochem. Photobiol. Sci.* **2015**, *14*, 1476–1491.
- (48) Cox, B. G. *Acids and Bases: Solvent Effects on Acid–Base Strength*, 1st ed.; Oxford University Press: Oxford, U.K., 2013; Chapter 7, pp 102–103.
- (49) Hurwitz, A. R.; Liu, S. T. *J. Pharm. Sci.* **1977**, *66*, 624–627.