

PREDICTING AQUEOUS pK_A

Predicting the pK_a of a compound from first principles remains a challenge, despite all of the many algorithmic and methodological advantages within the sphere of computational chemistry. Predicting the gas-phase deprotonation energy is relatively straightforward as I detail in Section 2.2 of my book. The difficulty is in treating the solvent and the interaction of the acid and its conjugate base in solution. Considering that we are most interested in acidities in water, a very polar solvent, the interactions between water and the conjugate base and the proton are likely to be large and important!

Baker and Pulay report a procedure for determining acidities with the aim of high throughput.¹ Thus, computational efficiency is a primary goal. Their approach is to compute the enthalpy change for deprotonation in solution using a continuum treatment and then employ a linear fit to predict the pK_a with the equation:

$$pK_a(c) = \alpha_c \Delta H + \beta_c$$

where c designates a class of compound, such as alcohol, carboxylic acid, amine, etc. Fitting constants α_c and β_c need to be found then for each unique class of compound, where the fitting is to experimental pK_a s in water. In their test suite,

they employed eleven anilines and amines, seven pyridines, nine alcohols and phenols, and seven carboxylic acids.

They test a number of different computational variants: (a) what functional to employ, (b) what basis set to use for optimizing structures, and (c) what basis set to use for the enthalpy computation. They opt to employ COSMO for treating the solvent and quickly reject the use of gas phase structures (and particularly use of geometries obtained with molecular mechanics). Their ultimate model is OLYP/6-311+G**//3-21G(d) with the COSMO solvation model. Mean deviation is less than 0.4 *pK* units. They do note that use of HF or PW91 provides similar small errors, but ultimately favor OLYP for its computational performance.

While this procedure offers some guidance for future computation of acidity, I find a couple of issues. First, it relies on fitted parameters for every class of compound. If one is interested in a new class, then one must develop the appropriate parameters – and the experimental values may not be available or perhaps an insufficient number of them are experimentally available. Second, the parameters cover-up a great deal of problematic computational sins, like the solvation energy of the proton, small basis sets, missing correlation energies, missing dispersion corrections etc.

A purist might hope for a computational algorithm that allows for systematic correction and improvement in the estimation of pK_a s. Further work needs to be done to meet this higher goal.

Source: <http://comporgchem.com/blog/?p=663>