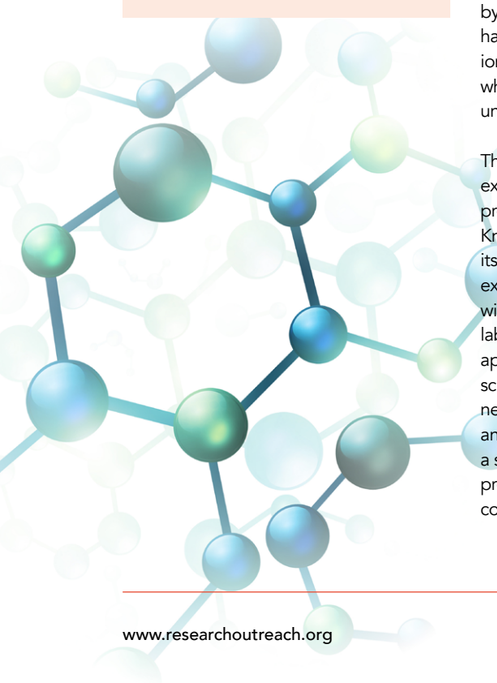
pK_a prediction from ab initio calculations

Most drugs, along with many other metabolically active compounds, behave in water solution as weak acids or bases. Their function and therapeutic activity are linked to their ability to exchange hydrogen ions with other molecules in physiological conditions. Professor Paul Popelier and Dr Beth Caine at the Manchester Institute of Biotechnology have developed a powerful and accurate method to predict the acid/base properties of prospective drugs using a simple and general criterion for linking quantum chemically-derived molecular geometry to acidity. Remarkably, the method has also been used to identify and correct for erroneously measured experimental values.



Acidity is one of the most important properties of many chemically active compounds. A compound behaves as an acid if it shows a tendency to transfer one or more hydrogen ions (H^+, or protons) to an acceptor species. In the context of biochemical processes in living organisms, often the acceptor species are water molecules. Dissolving an acidic molecule in water will increase the concentration of protons in solution. Vice versa, basic molecules accept protons from water, therefore reducing the H^+ concentration.

WEAK ACIDS AND BASES: THE PK_A

The chemical activity of biological molecules, metabolites and drugs is crucially related to their ability to exchange protons. The biological function of many molecules relies on their ability to carefully regulate how protons are exchanged with other chemical species. The ability of a weak acid to dissociate in water (i.e. donate protons) is measured by its pK_a, which is the pH at which half of the molecules in solution are ionised (their proton has dissociated), while the rest of the molecules remain unionised (undissociated).

The pK_a of a molecule can be measured experimentally, alternatively, it can be predicted using theoretical models. Knowing a drug's pK_a helps to predict its activity, distribution, metabolism, excretion and toxicity (ADMET) profile without having to synthesise it in the laboratory. In principle, *in silico* prediction approaches can drastically speed up the screening of large numbers of potential new drugs. Paul Popelier, Beth Caine and their collaborators have developed a simple and efficient approach to pK_a prediction, which uses information concerning the 3D structure of one protonation state of a molecule to model its behaviour as a weak acid or base.

MOLECULAR STRUCTURE AND ACIDITY

The relative acidity of compounds with similar chemical structures is determined by the electronic properties of other atoms that are present within the molecule (substituent groups). For instance, when a weak acid releases a proton, a negative charge is left behind and it is said to be ionised. The extent to which this charge is redistributed and stabilised throughout the molecule dictates the relative propensity of the compound towards ionisation. The Popelier group have shown that for a series of molecules with similar structures and quantum-chemical properties (electronic congeners), specific bond distances are strongly correlated to pK_a values. Consequently, after a model has been calibrated for a series of congeners using experimental data, the only information required to compute pK_a values for new compounds is bonding distances for a stable molecular geometry. Crucially, this criterion applies equally well to existing molecules and to new (never synthesised) species, provided a sufficient degree of structural similarity can be identified among them.

AB INITIO BOND LENGTHS PK_A PREDICTION

Stable, and therefore commonly observed molecular geometries can be determined using quantum-chemical calculations and a procedure called geometry optimisation. One of the most powerful and commonly used approaches to carry out geometry optimisations (and, in general, electronic structure calculations) is known as density-functional theory (DFT).

Gorodenkoff/Shutterstock.com

The pK_a of a drug helps to predict useful characteristics (its ADMET profile).

Prof Popelier and his collaborators have systematically studied the application of DFT to the calculation of pK_as based exclusively on structural descriptors (equilibrium bond lengths) for several classes of molecules. The team have developed a robust workflow that provides accurate estimates of pK_as using regression modelling. Their method, called *Ab Initio* Bond Lengths-pK_a (AIBL-pK_a) has been shown to be extremely reliable: not only can it be used to predict pK_as of existing or new molecules very accurately, it can also be applied to revise and indeed correct experimental estimates of pK_as when these are affected by measurement errors or inaccuracies.

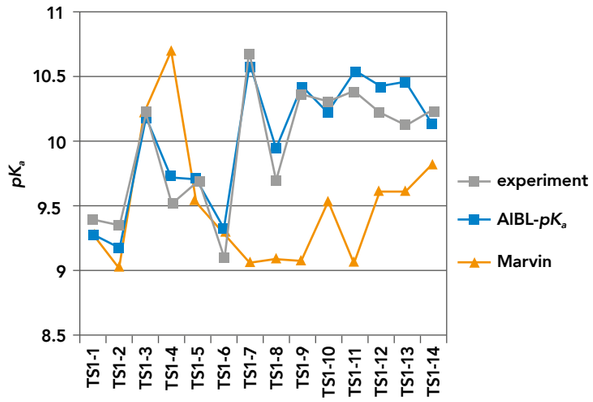
CORRELATION OF PK_A AND BOND LENGTHS

The AIBL-pK_a method has been applied successfully to several classes of organic molecules, some of which constitute the building blocks of more complex and biologically important compounds. For instance, in a study of 171 phenol-based molecules (chemical compounds containing an acidic –OH functional group linked to a conjugated phenyl ring), AIBL-pK_a has been shown to predict pK_as with an accuracy of less than 0.5 logarithmic units compared to experiments. A detailed statistical analysis of potential correlations between molecular structure and pK_a has also indicated the existence of a very strong correlation between a single chemical bond length in the molecule (calculated using DFT) and the acidity of the molecule. This is a striking and far reaching result, which shows that pK_a values can be estimated to very good accuracy using a single molecular reactivity index (a bond length). Furthermore, this study has demonstrated that the best accuracy in pK_a predictions can be achieved by splitting sets of analogue molecules into high correlation subsets (HCSs), which group together molecules with similar structural features, particularly in the close vicinity of the functional group that releases the proton.

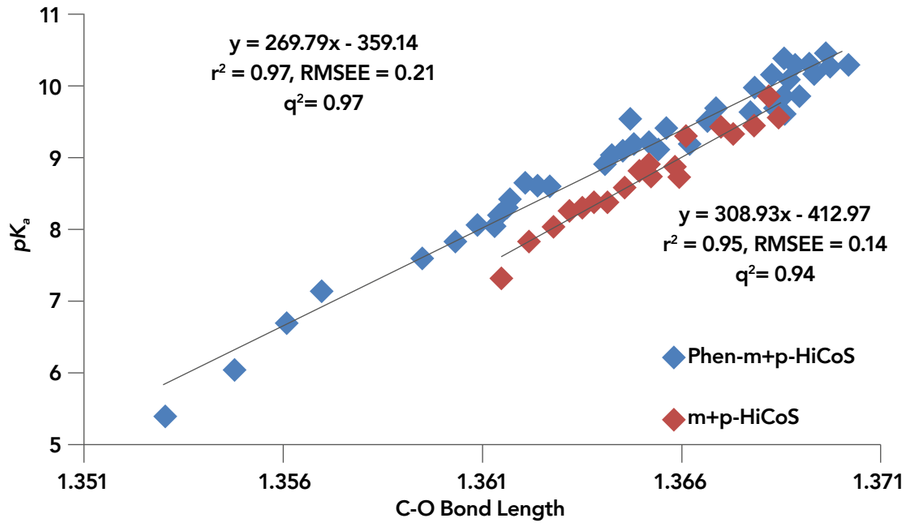
A GENERAL APPROACH TO PK_A CALCULATION

Other instances of where AIBL has been successful include: benzoic acids and anilines, carboxylic acids, amidine and guanidine-based compounds, primary and secondary sulfonamides

Knowing a drug's pK_a makes it possible to predict its activity, distribution, metabolism, excretion and toxicity.



Relative pK_a values of sulfonamide groups in 14 drugs. Although Marvin's performance in terms of mean absolute error evaluation is good (0.65), AIBL-pK_a manages to closely match the overall trend in the magnitude of pK_a values across the series.



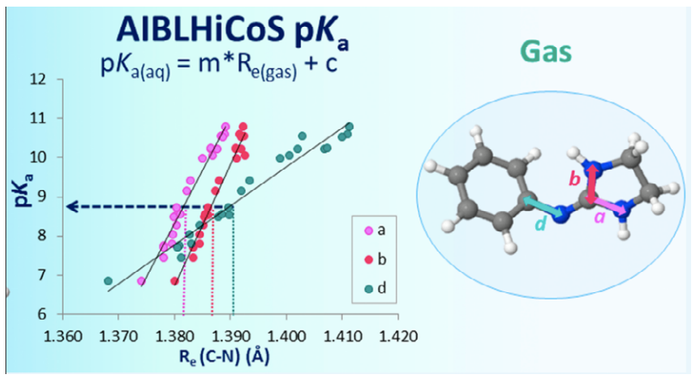
A simple linear relationship exists between the equilibrium bond lengths—that is, the distance between atoms for a stable arrangement of a molecule—and it's pK_a value, shown here for the C-O bond in phenol derivatives.

and a variety of carbon acids. In all cases, these studies have shown the existence of statistically significant bond length/pK_a correlations, along with the appearance of HCSs within groups of chemical analogues. The AIBL-pK_a method has also been highly successful in cases for which other methods for pK_a prediction have been shown to be unreliable or difficult to apply, including molecules containing more than one acidic functional group and systems that exhibit tautomerism, i.e. the coexistence of two rapidly interconverting chemical structures for the same molecule. Tautomerism is observed in several classes of organic and biological compounds, including amino acids (the fundamental constituents of proteins) and nucleic acids (the building blocks of DNA).

NOVEL APPLICATION OF THE AIBL-PK$_A$ APPROACH

Empirical predictors need experimental data to train their model(s). For regions of chemical space that are not properly represented in the training set, the accuracy of these methods rapidly deteriorates. A recent study carried out in collaboration with Lhasa Ltd. has shown that the AIBL approach can provide a solution to this problem. Using AIBL-pK_a models constructed for variants of carbon acids (a region of chemical space where Lhasa's empirical-based model performs least well), Professor Popelier and co-workers have constructed hypothetical compounds and predicted their pK_a values. These hypothetical compounds have been purposefully built to increase the diversity of atom types found in the carbon acids subset of the training set. Excitingly, their results indicate that the addition of such species to the training set both enhances the accuracy of Lhasa's predictive tool and makes it and more widely applicable. The current work of Prof Popelier and his collaborators aims to expand further and document the reliability and accuracy of the AIBL-pK_a method to systems of increasing complexity. Further work will explore how the approach can be applied to monoterpenes, and thus aid the rational design of new monoterpene synthase enzymes.

The team have developed a robust framework that provides accurate estimates of pK_a.

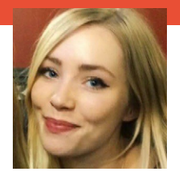


Highly correlated linear relationships exist between C-N bond lengths and pK_a in guanidine-containing compounds.

Behind the Research

Professor
Paul Popelier

E: paul.popelier@manchester.ac.uk
W: www.qct.manchester.ac.uk/



Beth Caine PhD

E: paul.popelier@manchester.ac.uk
W: www.qct.manchester.ac.uk/

Research Objectives

Research in Professor Popelier's laboratory focuses on next-generation force field design using machine learning, dynamic electron correlation and chemical insight from modern wave functions.

MANCHESTER 1824
The University of Manchester

Detail

Manchester Institute of Biotechnology (MIB)
Princess Street, Manchester
M1 7DN

Bio

Educated in Flanders to PhD level, Paul Popelier is a Professor of Chemical Theory and Computation at the MIB with more than 240 publications including 3 books. According to the Web of Science his contributions fall in 9 categories but are largely underpinned by quantum chemical topology and method development, including the novel force field FFLUX.

Beth Caine is a research scientist at the MIB, where she also earned her PhD. Her research aims to provide accurate models for acidity and basicity variation of organic compounds, aiding development of new pharmaceuticals and agrochemicals.

Funding
- BBSRC
- Syngenta
- Lhasa Ltd.

Collaborators
- Nathan Kidley
- Christophe Dardonville
- Maddalena Bronzato
- Torquil Frazer

Personal Response

One of the most striking findings of your work has been that a complex chemical phenomenon, the dissociation of a weak acid in water, can be rationalised using a simple criterion based on easily calculated structural information on one protonation state. What are the prospects of the AIBL-pK$_a$ method in drug discovery, and are there any challenges to its current applicability or potential improvements that it could benefit from?

The AIBL method can be useful in drug discovery in a number of contexts. During lead optimisation, experimental pK_a measurements are often taken for a series of synthesised analogues. Our method takes that information, plus quantum chemical information, and the result is a model that can not only accurately predict pK_a variation with structure, but also identify when experimental values have been measured erroneously. This correction of experiment has been shown numerous times, most recently in the case of marketed sulfonamide/sulfonylurea drugs called celecoxib, glimepiride and glipizide. Therefore, as well as making predictions on new compounds, the AIBL-pK_a approach can be used to check the consistency of a group of pK_a measurements, and thus serve as a rectifier for experimental outliers.

A caveat of AIBL is that it requires calibration, and each linear model has a defined domain of applicability that is restricted by the availability of experimental data. Furthermore, quantum chemical calculations are more time-consuming than other methods, such as Lhasa's approach. Their approach uses 2D molecular fingerprints to define a molecular structure (no quantum chemical input features) and returns predicted values in a matter of seconds. As we have mentioned, one exciting example of AIBL's potential is feeding a fast empirical model with theoretical, yet highly accurate data. This has already been shown to improve accuracy for Lhasa's approach with the addition of less than 150 hypothetical compounds, whilst still providing the user with an answer on a very short timescale.

References

Quantum Chemical Topology Group (2015). Quantum Chemical Topology, University of Manchester. www.qct.manchester.ac.uk/

Harding, AP., *et al.* (2011). pKa prediction from an ab initio bond length: part 2—phenols. *Phys. Chem. Chem. Phys.* 13, 11264-11282. doi:10.1039/C1CP20379G.

Harding, AP., *et al.* (2011). pKa prediction from an ab initio bond length: Part 3—benzoic acids and anilines. *Phys. Chem. Chem. Phys.* 13, 11283-11293. doi:10.1039/C1CP20380K.

Alkorta, I., *et al.* (2013). Relationship between experimental pKa values in aqueous solution and a gas phase bond length in bicyclo[2.2.2]octane and cubane carboxylic acids. *J. Phys. Org. Chem.*, 26, 791-796. doi:10.1002/poc.3159.

Griffiths, MZ., *et al.* (2013). Predicting pKa Values in Aqueous Solution for the Guanidine Functional Group from Gas Phase Ab Initio Bond Lengths. *Mol. Informatics*, 32, 363-376. https://doi.org/10.1002/minf.201300008.

Alkorta, I., *et al.* (2015). Linear Free-Energy Relationships between a Single Gas- Phase Ab Initio Equilibrium Bond Length and Experimental pKa Values in Aqueous Solution. *ChemPhysChem*, 16, 465-469. https://doi.org/10.1002/cphc.201402711.

Anstöter, C., *et al.* (2016). The AIBLHiCoS Method: Predicting Aqueous pKa Values from Gas- Phase Equilibrium Bond Lengths. *J. Chem. Inform. Modeling*, 56, 471-483. https://doi.org/10.1021/acs.jcim.5b00580.

Dardonville, C., *et al.* (2017). Substituent effects on the basicity (pKa) of aryl guanidines and 2-(arylimino)imidazolidines: correlations of pH-metric and UV-metric values with predictions from gas-phase ab initio bond lengths. *New J. Chem.* 41, 11016-11028. doi:10.1039/C7NJ02497E.

Caine, BA., *et al.* (2018). Prediction of Aqueous pKa Values for Guanidine-Containing Compounds Using Ab Initio Gas-Phase Equilibrium Bond Lengths. *ACS Omega*, 3, 3835-3850. https://doi.org/10.1021/acsomega.8b00142

Caine, BA., *et al.* (2019). Experiment stands corrected: accurate prediction of the aqueous pKa values of sulfonamide drugs using equilibrium bond lengths. *Chem. Sci.* 10, 6368-6381. doi:10.1039/c9sc01818b. doi:10.1039/C9SC01818B.