Prof. John Blacker and Dr Bao N. Nguyen

j.blacker@leeds.ac.uk; b.nguyen@leeds.ac.uk
phone: 0113 343 8239; 0113 343 0109

This proposal is representative of the projects currently on offer in our groups. For more details of active research projects, please visit the links: http://www.chem.leeds.ac.uk/john-blacker.html; http://www.chem.leeds.ac.uk/bao-nguyen/research-summary.html

Solubility prediction in organic solvents through a combination of chemometrics and computational chemistry

Solubility is an essential property in evaluating an Active Pharmaceutical Ingredient's potency. In pharmaceutical syntheses, solubility is vital in assisting purification of intermediates and synthetic route selection by predicting the cost of work-up and purification. Consequently, significant effort has been expended in developing reliable and high throughput techniques to measure solubility experimentally.[1] However, these require large quantities of pure compounds which are often unavailable.

Most current models rely on experimental data, either as thermodynamic values, or as parameters for structural fragments. These allow semi-empirical models to adjust and compensate for inherent errors in their assumptions. Problems, however, arise when novel compounds are made with completely different structural patterns to those used to provide the semi-empirical parameters, *i.e.* outside the known chemical space.

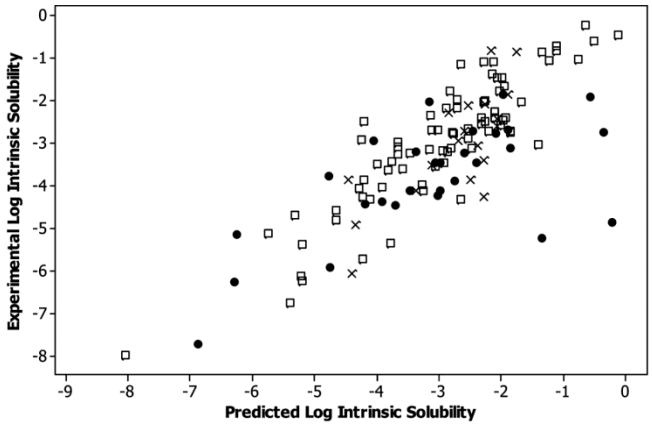

Figure 1. Experimental *vs* predicted logS (solubility) using a solubility model for a test set of compounds (•), using a training set of data (□), the validation set of data (x). ***The right hand side points are particularly poor predictions***.[2]

The student will take a leading role in addressing this urgent gap in chemical knowledge. The project will exploit recent advances of Density Functional Theory (DFT) computational techniques to calculate properties of drug-like compounds. Statistical tools will be employed to analyse these properties and link them with experimental solubility.[3] A model for solubility prediction will consequently be developed, with understanding of factors influencing solubility. Importantly, the DFT approach is not restricted to known chemical space. This will allow the project to quickly expand into novel chemical space, and provide predictions for experimental verification in the later stage.

The project will be carried out in collaboration with fine chemical/pharmaceutical companies to fully understand and to address their needs. The predictions will be demonstrated on real pharmaceutical compounds.

The project is best suited to a student in Chemical Engineering or Chemistry with strong background and interest in chemometrics and computational chemistry. Additional training on programing languages, *e.g.* python, R, statistics and experimental solubility measurements will be provided. These are important transferable skills in both academia and industry. The student will also benefit from interdisciplinary training and seminar programmes in process chemistry as a member of the Institute of Process Research & Development, Leeds (http://www.iprd.leeds.ac.uk/).

A 3-months placement at AstraZeneca to transfer the results of the project and take advantages of high throughput facilities is an integral part of the project.

More detail on this and other projects in asymmetric catalysis, recovery of precious metals, or CO_2 utilisation will be made available by contacting Prof. John Blacker at j.blacker@leeds.ac.uk and Dr Bao N. Nguyen at b.nguyen@leeds.ac.uk.

Funding Status: Funded Project (UK/EU students)

References

(1) J. Alsenz, *Adv. Drug Deliv. Rev.* **2007**, *59*, 546; (2) J. Wang, *Comb. Chem. High T. Scr.* **2011**, *14*, 328; (3) for a similar approach in predicting ligands for catalysis, see N. Fey, *Organometallics* **2010**, *29*, 6245; N. Fey, *Organometallics* **2008**, *27*, 1372; N. Fey, *J. Chem. Inf. Model.* **2006**, *46*, 2951;