MS29 O4

True Prediction of H-bonding Interactions: A Knowledge-based Model of Hydrogen-Bonding Propensity in Organic Crystals Peter T. A. Galek,[a,b] László Fábián[a,b] and Frank H. Allen[a] [a]*Cambridge Crystallographic Data Centre, Cambridge, UK.* [b]*Pfizer Institute for Pharmaceutical Materials Science, Department of Engineering, University of Cambridge, UK.* E-mail: galek@ccdc.cam.ac.uk

Keywords: hydrogen bonding, nonbonded interactions, logistic regression, Cambridge Structural Database

A new method is presented to predict which hydrogen donors and acceptors form hydrogen bonds in a crystal structure, based on statistical analysis of hydrogen bonds in organic molecules in the Cambridge Structural Database (CSD) [1]. Descriptive molecular and chemical properties are accumulated using new survey software, which are subsequently used to create a dichotomous probability model to estimate the likelihood of H-bond formation between donor D and acceptor A.

The logistic regression technique is applied assuming the logit probability function. The model discriminates according to the value and type of parameter; thus the *latent* information in the CSD is accessed collectively to predict specific structure directing interactions in organic crystals. The result is a pair-wise propensity to form an H-bond; the complete description has been termed the *Logit Hydrogen-bonding Propensity* (LHP) model.

Initial applications will be demonstrated showing the LHP model to be highly accurate, achieving ~90% correct classification of observed pairs in the dataset (both observed H-bonds, and non-interacting donor-acceptor pairs). Specific issues with the model will be presented also. *Holdout validation* (splitting the data into two sets) measures model flexibility and any dependence on particular fitting data, using a model fitted with a training set and calculating propensity scores for a validation set. Indeed, the validation data may form a true "blind test" for the model. Initial tests will be discussed showing the model to be remarkably robust across a range of structures.

Interesting insights are obtained by separating crystal structures by age. For example, propensities for crystal structures archived to the CSD post-1995 have been calculated using a model built with pre-1995 crystal structures alone. Examples will be presented using various datasets of related chemical functionality. In the proper sense, such examples are true predictions.

[1] F. H. Allen (2002). *Acta Cryst.*, B58, 380-388.

MS29 O5

Predicting Hydrogen-Bond Strengths by Structural and Thermodynamic Databases Paola Gilli and Loretta Pretto, *Chemistry Department and Centre for Structural Diffractometry, University of Ferrara, Ferrara, Italy.* E-mail: paola.gilli@unife.it

Keywords: H-bond strength, thermodynamic databases, structural databases

The *PA/pKₐ equalization principle* [1,2], that is the idea that H-bond strengths depend on the matching of the donor and acceptor acid-base properties, is more than forty years old but has never been verified in its generality for lack of suitable data. Its definitive verification is now possible, at least in principle, because of the availability of ever more extended compilations of thermodynamic acid-base indicators (proton affinities, PAs, in the gas phase [3] and $pK_a(H_2O)$'s in aqueous solutions [4]) and of two comprehensive sources of H-bond strengths (the NIST database, reporting gas-phase dissociation energies for hundreds of H-bonded complexes [3], and the Cambridge Structural Database, CSD, including the crystal structures and H-bond geometries of thousands of chemical compounds [5,6]).

The combination of two acid-base indicators (PA and pK_a) with two datasets (gas-phase energies and solid-state geometries) provides four possible correlations between H-bond strengths and donor-acceptor acid-base properties, each one potentially leading to the full assessment of the equalization principle. This does not occur in practice because a number of constraints of thermodynamic nature severely limit the application of each type of correlation to partial subsets (classes) of H-bonds and, sometimes, to none.

The main practical consequences are that: *(i)* resonance-assisted H-bonds (RAHBs) cannot be interpreted by means of any acid-base indicator; *(ii)* gas-phase dissociation enthalpies can be used for the H-bonds which are charged associations (CAHBs) of two acids, [A..H..A'] or two bases, [B..H..B']$^+$, but not for those which are acid-base associations, A-H…B, and constitute the by far most numerous class of H-bonds; and *(iii)* only correlations between H-bond strengths deriving from crystal geometries and pK_a values, *i.e.* pK_a differences between H-bond donor and acceptor groups, can deal with these three classes of bonds at the same time and have, therefore, the chance to become the method of election for a quick and reasonably accurate prediction of H-bond strengths based on thermodynamic acid-base indicators.

The present communication, beside discussing the more theoretical thermodynamic aspects of the problem, will treat in detail the difficulties encountered in the practical comparison of acid-base parameters with geometrical and thermodynamic H-bond-strength indicators. The final aim is to develop an easy computer-based procedure able to provide reasonable energy estimates for all the H-bonds whose geometries become known from structural crystallography studies.

[1] Gilli P., Bertolasi V., Pretto L., Ferretti V., Gilli G., *J. Am. Chem. Soc.*, 2004, 126, 3845.
[2] Gilli P., Bertolasi V., Pretto L., Antonov L., Gilli G., *J. Am. Chem. Soc.*, 2005, 127, 4943.
[3] Linstrom P.J., Mallard W.G. Eds., NIST Chemistry WebBook, June 2005 (http://webbook.nist.gov/chemistry/).
[4] Martell A.E., Smith R.M., Motekaitis R.J., NIST Critically Selected Stability Constants of Metal Complexes, NIST, May 2004.
[5] Allen F.H., Bellard S., Brice M.D., Cartwright B.A., Doubleday A., Higgs H., Hummelink T., Hummelink-Peters B.G., Kennard O., Motherwell W.D.S., Rodgers J., Watson D.G., *Acta Cryst.*, 1979, B35, 2331.
[6] Allen F.H., *Acta Cryst.*, 2002, B58, 380.