Systematic Study into the Salt Formation of Functionalised Organic Substrates

S. C. Ward[a], M. B. Hursthouse[a], D. C. Woods[b] and S. M. Lewis[b]

a School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK.
b School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK.

The objective of this ongoing work is to perform a detailed systematic study of organic salt formation through a series of designed experiments. We have identified a set of descriptors that describe molecular properties relevant to salt formation. For the initial experiments, a collection of salt forming acids have been assembled using the Cambridge Structural Database and their descriptor values have been calculated. These acids define a chemical space from which the compounds for the first experiments can be chosen. The experiments aim to explore this chemical space whilst building statistical models that will allow a better understanding of how the descriptors affect salt formation.

INTRODUCTION

There is currently a great amount of interest in the use of salts in the pharmaceutical industry because certain properties of the solid forms can be modified without altering the desired effect of the drug.

Salt formation is essentially a three component system involving an acid (A), a base (B) and one or more solvents. A salt is formed by the transfer of a proton (H^+) from an acid (A) to a base (B):

$$A\text{-}H + B \rightarrow (A^-)(B^+\text{-}H)$$

The majority of drugs are basic (B) and therefore a large proportion of the work involves selecting a suitable acid former.

Each salt imparts unique properties onto the parent compound. The selection of the best salt form for the ionisable drug is now of paramount importance in the pharmaceutical development of new chemical entities[1].

Typically, the first step in a salt selection procedure is the formation of a wide variety of salts, followed by the selection of the most crystalline salt form produced. In order to assist salt selection a number of empirical rules have been devised, such as the 'rule of three'. This states that salt formation generally requires a difference of at least three pK_a units between the conjugate base and the conjugate acid,

$$pK_a \text{ (base)} - pK_a \text{ (acid)} \geq 3$$

where pK_a is the ability of an ionisable group to donate a proton (H^+) in an aqueous medium and is often referred to as the dissociation constant.

Although rules such as the one highlighted above are valuable guidelines, as far as we are aware no detailed study has been made of the reliability and/or basis of these empirical procedures. A carefully planned set of experiments may lead to a more scientific method for assessing the viability of salt formation, rather than relying on trial and error.

INVESTIGATING THE CAMBRIDGE STRUCTURAL DATABASE

An investigation of the Cambridge Structural Database (CSD)[2] was initially performed to identify acids which form salts and co-crystals. The objective was to identify when a salt forms in preference to a co-crystal. We consider succinic acid as an example to illustrate the process.

Succinic acid is a pharmaceutically acceptable acid and can exist in two polymorphic forms[3]. In total, five salts and eight co-crystals were found from the CSD, using specific criteria (Table 1). A co-crystal is taken to be an A-B composite in which no proton transfer has occurred:

$$A\text{-}H + B \rightarrow (A\text{-}H)(B)$$

Table 1 also gives the pK1[4] value, the pK_a for the most ionisable site in the molecule, for succinic acid and the salt and co-crystal forming bases. When a salt was formed, the differences in pK_a values between the acid and base were greater than 2.7 and so are in accordance with the approximate 'rule of three'. For the co-crystals, the differences between pKa values were generally less than for the salts and it is likely that the pKa values for the bases were not high enough to allow proton transfer.

The example given agrees with the hypothesis that a large difference in pK_a values leads to a higher chance of salt formation. However, other factors need to be considered and this is why characterising salt formation space is essential in the prediction of salt formation.

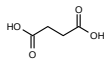
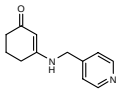
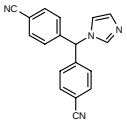
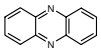
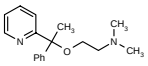
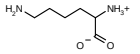
	Compound	Diagram	pK1
Acid	Succinic Acid		4.2 A
Co-crystal forming bases	Benzamide		*
	2-pyridone		0.7 B
	3-(4-picolinylamino) cyclohex-2-enone		5.7 B
	aminopyrimidine		3.9 B
	(p-cyanophenyl)imidazolylmethane		6.1 B
	Urea		0.1 B
	Phenazine		1.6 B
	2-amino-6-ethyl-4(3H)-pyrpyrimidone		0.5 B
Salt forming bases	piperazine		9.8 B
	Doxylamine		8.7 B
	L-Lysine		9.5 B
	Imidazole		6.9 B
	Ethylene-1,2-diammonium		9.9 B

Table 1. Co-crystals and salts of succinic acid with their corresponding pK1 values. A: indicates an acidic pK_a. B: indicates a basic pK_a. * pK1 of this compound has not been reported in the literature.

Figures 1 and 2 show the crystal structures for one of the salts and one of the co-crystals as an example.

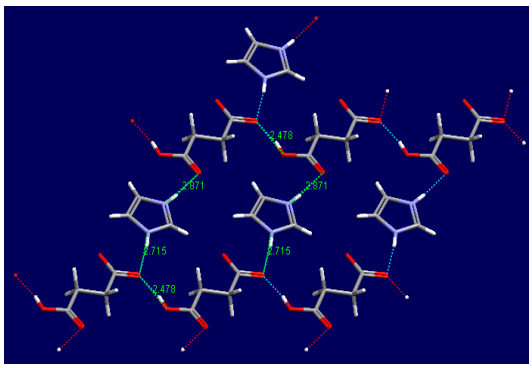


Figure 1. A packing diagram for imidazolium hydrogen succinate viewed along the *c* axis[5].

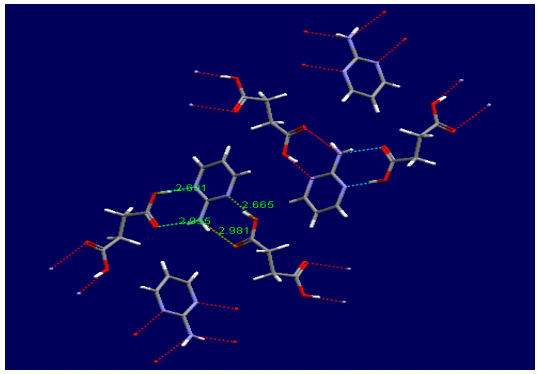


Figure 2. A packing diagram for 2-aminopyrimidine succinic acid, viewed along the *a* axis[6].

MODEL-BASED APPROACH

We are investigating an approach in which a statistical model, called a *response surface model*[7], is fitted to the data from a designed experiment. The fitted model may then be used to predict the combinations of acids and bases that are likely to produce a salt.

Due to the fact that there is a wide variety of choices for the acid or the base, a set of chemical descriptors was sought that could be used to characterise the chemical space of interest and to form a statistical model. The chosen descriptors should represent key aspects of the molecular structure, which relate to its salt forming ability. It is also preferable to have a diverse set of values for each descriptor in order to provide a wide choice of possible compounds that could be chosen for the experiments.

There were two main choices of descriptors:

- Traditional molecular descriptors – which are directly interpretable as properties of the molecule.
- BCUT (Burden Chemical abstracts service University of Texas)[8] type descriptors – which are single number descriptors that summarise the information in the molecular structure and the atoms in the structure, via eigenvalues of weighted connectivity matrices.

It was decided to investigate meaningful molecular descriptors that, from chemical knowledge, were considered most likely to be related to salt formation. A shortlist of such descriptors was eventually chosen that were tabulated in the literature or easily calculated.

DESCRIPTORS

As a starting point, an initial set of 67 acids was obtained using the CSD. The selected descriptors were either found in the literature or calculated using software such as HyperChem[9]. Values for a total of ten descriptors were investigated. A particular concern was to avoid the use of descriptors which are strongly related, for example including pairs of descriptors which are highly correlated. These may lead to redundant terms in the fitted model and coefficient estimators which are difficult to interpret and have high standard errors.

Figure 3 shows a matrix of plots of all the two-dimensional projections (scatter plots) of the values of the ten descriptors (labelled X_1 to X_{10} for simplicity) for the acids. These scatter plots show the relationship between pairs of descriptors for the available acids. A high proportion of points along the diagonal indicates a strong correlation between two descriptors. Prior to investigating the descriptor values, it was expected from chemical knowledge that several of the descriptors would be related and that those of most interest would be X_1, X_2, X_3 and X_5.

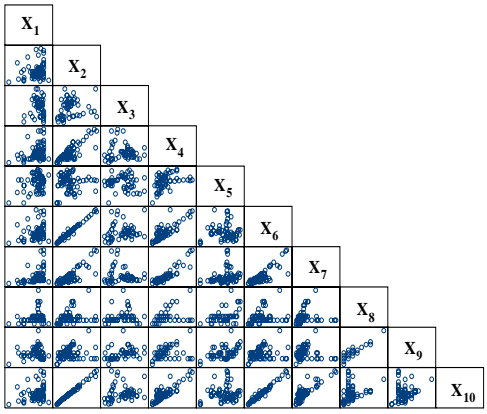


Figure 3. Two-dimensional projections for $X_1 - X_{10}$.

From the projections it can be seen that X_2, X_6 and X_{10} are strongly related to each other. There are also high (greater than 0.8) positive correlations between the pairs (X_2, X_4), (X_2, X_7), (X_4, X_7), (X_6, X_7), (X_7, X_{10}), (X_4, X_7) and (X_8, X_9). The removal of X_6 and X_{10} from the list of descriptors still leaves four highly correlated pairs. If X_4 and X_7 are removed, and only one of X_8 or X_9 (which are chemically closely related) is retained in the set, then the remaining variables appear to be unrelated. This results in descriptors X_1, X_2, X_3, X_5, together with either X_8 or X_9.

Table 2 shows the correlation matrix for the chosen descriptors, using X_8 rather than X_9. It should be noted that, when the descriptor comparisons were made, some acids had missing values for one or more descriptors. For the purposes of examining the correlations between pairs of descriptors, only those compounds for which values of both descriptors were available were used. Other, more sophisticated, approaches to descriptor selection could also be used but are not discussed here.

	X_1	X_2	X_3	X_5	X_8
X_1	1.000	0.246	-0.003	0.331	0.164
X_2	0.246	1.000	0.167	0.049	0.252
X_3	-0.003	0.167	1.000	-0.065	0.152
X_5	0.331	0.049	0.065	1.000	0.453
X_8	0.164	0.252	0.152	0.453	1.000

Table 2. Correlation matrix for the descriptors X_1, X_2, X_3, X_5 and X_8.

The next step in the process was to extend the set of acids to obtain better coverage of salt formation space. This was achieved by first identifying regions in the descriptor space where acids were sparse and then finding additional acids in these regions.

Figure 4 shows the two-dimensional projections of the chosen descriptors (X_1, X_2, X_3, X_5 and X_8).

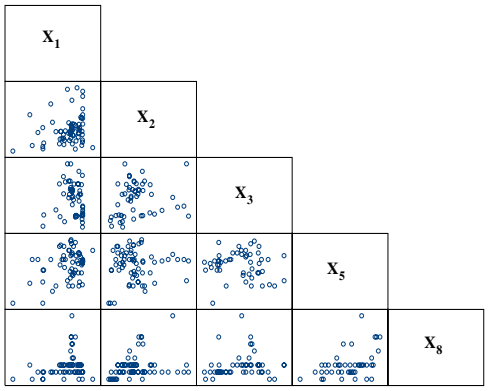


Figure 4. Two dimensional projections for X_1, X_2, X_3, X_5 and X_8.

These projections indicate a reasonable coverage, with the poorest coverage occurring in the X_1, X_3 projection (this is partly due to a number of missing values for the X_3 descriptor). From Figure 4 and the three-dimensional projections (not shown), it was decided to try to find additional compounds with either low X_1 values, high X_2 values or low X_5 values (or, ideally, combinations of these). The ranges of descriptor

values covered were also carefully considered to ensure they were appropriate for the initial experiments.

A total of 36 additional acids were added to the original set and their corresponding descriptor values obtained or calculated. The next step of the process is the careful choice of compounds for the initial experiments from the acids.

EXPERIMENTAL DESIGN

When making a selection of compounds from a chemical space for experimentation, it is often required to choose a subset that is either as diverse or as representative of the space as is possible[10]. To achieve these aims, it is common to use either *spread* or *coverage* designs. A spread design aims to have the selected compounds as spread-out as possible in the chemical space, whereas a coverage design ensures that each unselected compound is as close to a selected compound as possible. In the ongoing work, model-based design will also be considered, where the design aims to enable predictions to be made from the model as accurately as possible.

For our initial set of 67 compounds, and using the descriptors X_1, X_2, X_3, X_5 and X_8, Figure 5 shows the two-dimensional projections of a 24 point coverage design. The points in the coverage design are evenly spread across the possible compounds, giving similar, but less dense, projections compared with Figure 4.

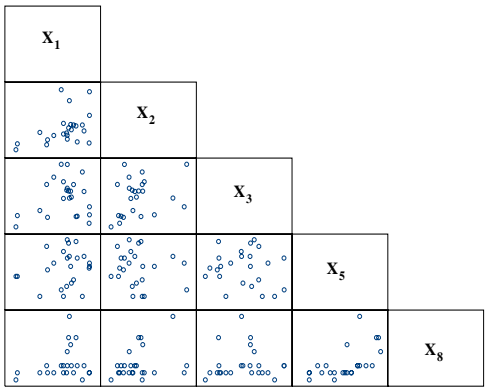


Figure 5. Two-dimensional projections for a coverage design for five descriptors.

CONCLUSION

A set of descriptors for investigating salt formation has been identified. These descriptors can now be used in experiments to investigate the properties needed for salt formation to occur. The eventual aim is that a database containing the descriptor values will be available to the scientific community over the Grid, together with rules that indicate which counter-ion would be most appropriate for a drug with a specified set of descriptors. It is also planned to make available search algorithms for finding suitable designs for the experiments via a software node on the Grid.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the EPSRC e-Science programme (GR/R67729, Combechem) along with AstraZeneca in Mölndal, Sweden.

REFERENCES

1. S. M. Berge, L. D. Bighley, and D. C. Monkhouse, *J. Pharm. Sci.,* 1977, **66**, 1.
2. F. H. Allen, *Acta Crystallogr. Sect. B*, 2002, **58**, 380-388.
3. N. N. Petropavlov, S. B. Yarantsev, *Kristallografiya,* 1983, **28**, 1132.
4. D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution: Supplement*. 1972, London: Butterworths Scientific Publications.
5. J. C. MacDonald, P. C. Dorrestein and M.M. Pilley, *Cryst. Growth Des*., 2001, **1**, 29- 38.
6 M. C. Etter, D. A. Admond and Doyle Britton, *Acta Cryst*., 1990, **C46**, 933-934.
7. R. H. Myers and D. C. Montgomery, *Response Surface Methodology* (2nd ed.), 2002, New York: Wiley.
8. R. S. Pearlman and K. M. Smith, *Perspect. Drug Discov*., 1998, **9**, 339-353.
9. HyperChem, Inc. 115 NW 4th Street, Gainsville, Florida 32601, USA.
10. R. E. Higgs, K. G. Bemis, I. A. Watson and J. H. Wikel, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 861-870.