Chemical Database Service (CDS) NEWSLETTER

Issue 22. Oct 2005

CCLRC Daresbury Laboratory

See http://cds.dl.ac.uk/cds/newsletters.html for this and previous editions of our newsletter

A summary of the main articles

The CDS Road Show programme for 2005/06 is presented on page 2.

Two articles from our research highlights are presented on page 3-6:

(a) The use of polymer supported isoureas for the *O*-alkylation of carboxylic acids

(b) Links between the CrystalWeb & ICSD-WWW database interfaces.

See page 7 for how CDS might look in the future.

For how to acknowledge the service, information on new data as well as the new features in the Cambridge Mercury package see Page 8.

EPSRC

The CDS is funded by the Engineering and Physical Sciences Research Council

Inside this issue:

CDS Road Show 2005/06

After the success of our 2004/05 roadshow programme, the CDS have provisionally identified 12 universities to visit on our forthcoming roadshow programme. The roadshow will take the same format as before (i.e. poster display/registration in departmental concourse in the morning, followed by a CDS overview presentation in the afternoon).

During our visit to Biological Sciences we will be talking to as many researchers as possible about what resources they use or would like to have access to. The overall aim is to improve our offering to the bioscience community and better understand the information needs of the biosciences community.

Visit	Site	Date
1	East Anglia Chemistry	29/09/05
2	Birmingham Chem. Eng	06/10/05
3	Nottingham Pharmacy	20/10/05
4	Newcastle Chemistry	27/10/05
5	Manchester Chem. Eng.	03/11/05
6	Liverpool John Moores Biosci.	10/11/05
7	Aberdeen Med. & Bio. science	17/11/05
8	Cambridge Chem. Eng.	24/11/05
9	London School of Pharmacy	26/01/06
10	Sunderland Bioscience	01/02/06
11	Southampton Chemistry	09/02/06

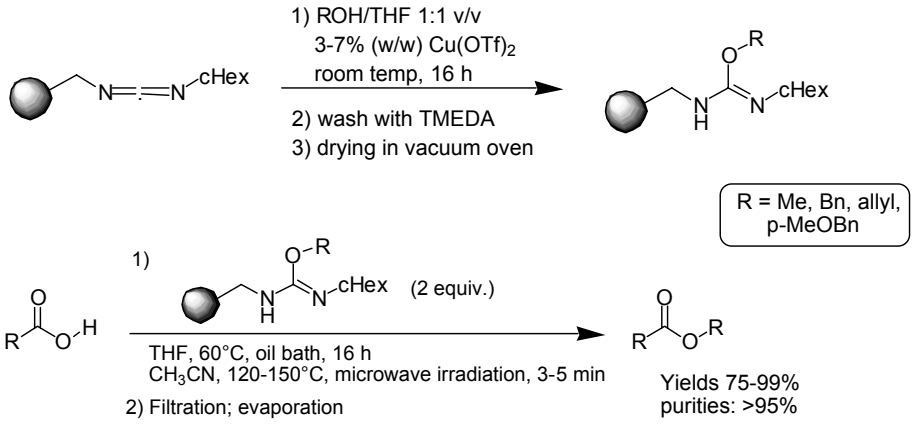
The use of polymer supported isoureas for the *O*-alkylation of carboxylic acids

Dr Bruno Linclau, School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ
bruno.linclau@soton.ac.uk

The breakthrough of combinatorial chemistry was made possible by the development of solid supported reactions. By immobilising the substrate onto an insoluble polystyrene support, the reaction workup could be drastically simplified as a simple filtration could be employed to remove excess reagents, byproducts etc. However, this method also had significant disadvantages, such as difficult reaction monitoring and the fact that unreacted substrate or side-products could not be separated from the product as all were linked to the insoluble support.

In polymer-assisted-solution-phase (PASP) chemistry, the *reagents* (or catalysts) are immobilised onto an insoluble support, while the substrate and products remain in the solution phase. Equally, reaction workup consists of a simple filtration, with the desired product now located in the filtrate. Obviously, for optimal efficiency, the system has to be designed such that all reagent byproducts remain linked to the support. This methodology combines the main advantages of traditional solution phase chemistry (eg reaction monitoring) with those of solid-supported chemistry (filtration as reaction workup), allowing for significant gains in time-efficiency. Another advantage is that there are no restrictions any more for the substrates used in terms of a necessary functional groups for linking to an insoluble support. Because the product is not immobilised, should the need arise, a purification can still be performed in case unreacted starting material remains, or side products are formed. This purification could be effected by polymer supported scavengers, or by traditional chromatography. PASP is currently a much used method in the pharmaceutical industry for the synthesis of small libraries of compounds. The use of PASP chemistry in multisptep synthesis has also been established. Hence, the development of new solid supported reagents suitable for PASP methodology is very important.

The synthesis of esters from carboxylic acids is a very common transformation in organic synthesis, yet, efficient PASP methodology was not available until a few years ago. We have successfully investigated the use of solid-supported isoureas for the *O*-alkylation of carboxylic acids. Initially we studied the protection of carboxylic acids using simple alcohols. The required solid-supported isoureas could easily be synthesised from solid-supported carbodiimide by reaction with an excess of alcohol and a copper (II) catalyst. Solid-supported carbodiimides are commercially available, but we also have developed an improved synthesis for immobilised carbodiimides from aminomethyl polystyrene resin. It was found that simply heating an acetonitrile solution of the carboxylic acid with the immobilised isourea, followed by filtration/resin washing and evaporation of the solvent, afforded the ester products in good yield (typically >85%) with excellent purity (typically >95%). When microwave irradiation was employed, the actual reaction time was as short as 3-5 minutes! For the synthesis of methyl esters, our methodology clearly is a better approach than the use of diazomethane.



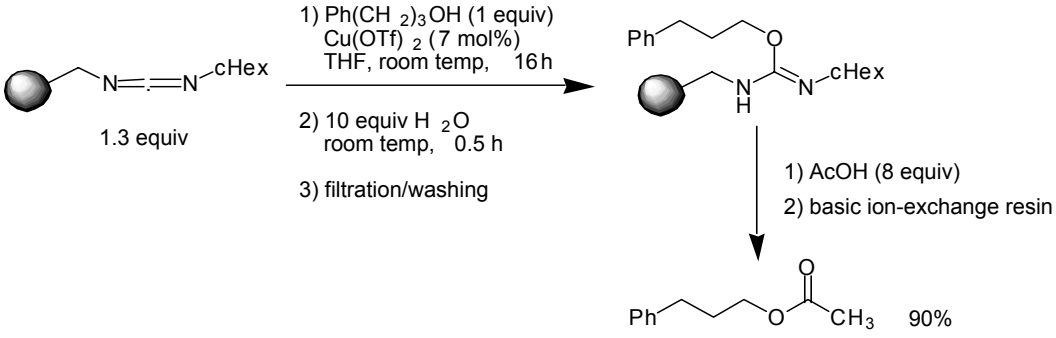
Continued on page 4

The use of polymer supported isoureas for the *O*-alkylation of carboxylic acids

Continued from page 3

The ester formation using more complex alcohols using a catch/release method is also under investigation. Catch/release is a 2-step process where the desired substrate switches phase twice, allowing for an extra purification step. Hence, the alcohol (1equiv) was attached to the resin as the corresponding isourea, which could be purified by filtration to remove unreacted alcohol and Cu (II) catalyst. Residual carbodiimide groups were inactivated by reaction with water. The purified isourea was then reacted with a carboxylic acid, "releasing" the alcohol moiety to the solution phase again, as the ester. Excess or unreacted carboxylic acid was removed by adding an immobilised basic scavenger immediately to the reaction mixture. Finally, a filtration removed all resin material, leaving pure ester

Polymer-assisted-solution-phase chemistry is a very active research field, and the use of reaction databases has proved to be an excellent tool to find necessary information. In particular, the 'Solid-Phase Synthesis' database contained in ISIS/Base, maintained by CDS, was found to be of excellent value to search for reactions on solid support. In addition, the ACD finder database also was very useful to quickly obtain information about available starting materials and substrates. The way the use of these databases has advanced our research is an illustration of the excellent value of the CDS service for the UK chemical community.



1. Review: Kirschning, A.; Monenschein, H.; Wittenberg, R. *Angew. Chem. Int. Ed.* **2001**, *40*, 651

2. Review: Clapham, B.; Reger, T. S.; Janda, K. D. *Tetrahedron* **2001**, *57*, 4637.

3. Review: Eames, J.; Watkinson, M. *Eur. J. Org. Chem*. **2001**, 1213.

4. Example: Storer, R. I.; Takemoto, T.; Jackson, P. J.; Ley, S. V. *Angew. Chem. Int. Ed.* **2003**, *42*, 2521.

5.(a) Adamczyk, M.; Fishpaugh, J. R.; Mattingly, P. G. *Tetrahedron Lett*. **1995**, *36*, 8345. (b) Pilot, C.; Dahmen, S.; Lauterwasser, H.; Bräse, S. *Tetrahedron Lett*. **2001**, *42*, 9179. (c) Rademann, J.; Smerdka, J.; Jung, G.; Grosche, P.; Schmid, D. *Angew. Chem. Int. Ed*. **2001**, *40*, 381. (d) Zander, N.; Gerhardt, J.; Frank, R. *Tetrahedron Lett*. **2003**, *44*, 6557. (e) Erb, B.; Kucma, J.-P.; Mourey, S.; Struber, F. *Chem. Eur. J.* **2003**, *9*, 2582. (f) Crosignani, S.; Gonzalez, J.; Swinnen, D. *Org. Lett*. **2004**, *6*, 4579. (g) Petricci, E.; Mugnaini, C.; Radi, M.; Corelli, F.; Botta, M. *J. Org. Chem*. **2004**, *69*, 7880. (h) Zhang, M.; Vedantham, P.; Flynn, D. L.; Hanson, P. R. *J. Org. Chem*. **2004**, *69*, 8340.

6. (a) Crosignani, S.; White, P. D.; Linclau, B.*Org. Lett*. **2002**, *4*, 1035-1037. (b) Crosignani, S.; White, P. D.; Linclau, B. *Org. Lett*. **2002**, *4*, 2961-2963. (c) Crosignani, S.; White, P. D.; Steinauer, R.; Linclau, B. *Org. Lett.* **2003**, *5*, 853-856. (d) Crosignani, S.; White, P. D.; Linclau, B. *J. Org. Chem.* **2004**, *69*, 5897-5905

Links between the CrystalWeb & ICSD-WWW database interfaces –

Bob McMeeking & Dave Fletcher (r.f.mcmeeking@dl.ac.uk & d.a.fletcher@dl.ac.uk)
The EPSRC Chemical Database Service at Daresbury Laboratory

Providing ready access to major crystallographic data collections has always been a key task for the Chemical Database Service (CDS) [1], and to meet this requirement the Service has developed the CrystalWeb system over the last several years. CrystalWeb provides "one stop shop" web browser access to all our crystallographic data [2]. The aim has been to make such data readily useable by experts in a wide range of field without the requirement of any special crystallographic expertise [3].

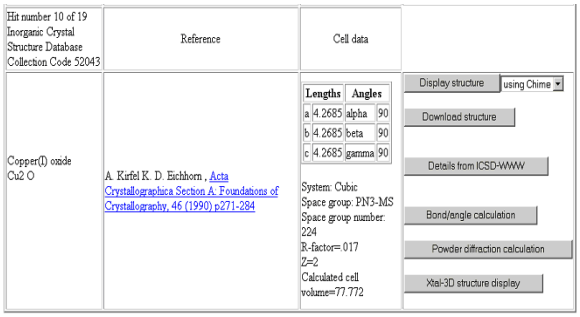
At the same time other established modes of access continue to be supported. In particular we grant access to the Cambridge Structural Database (CSD) of organics and organo-metallics via the ConQuest software and the Inorganic Structural Database (ICSD) using the ICSD-for-WWW developed by Alan Hewat and co-workers at the ILL Grenoble [4]. Like CrystalWeb ICSD-WWW is a web browser interface. However, it supports many valuable complementary features, having been built with inorganic structures specifically in mind.

A valuable option available in ICSD-WWW is having both atom (including thermal ellipsoid display) and polyhedron modes for its 3-D visualisation module [5]. There is also an interatomic distance calculator and an option to simulate and displaying neutron and X-ray powder diffraction patterns. The underlying calculations to generate the patterns are done by the program, Lazy, which uses Klaus Yvon *et al*'s well established Lazy Pulverix code. The user, however, is shielded from most low level details, and all these facilities are accessed via easy-to-use web browser forms. In almost all cases ICSD-WWW can be used as a "black box" with no need for user manipulation of the retrieved entries.

Historically a problem remained for the relatively few entries where the crystal structure was reported in a "non-standard" crystallographic setting (e.g. A2/m or I2/m rather than C2/m for Space Group number 12). These were handled automatically by CrystalWeb from its inception, but traditionally have required manual intervention when using ICSD-WWW. Problem entries are flagged, and in most instances the user needed to find and add the correct space group operators in the appropriate format. In other cases it was necessary to apply a translational correction to the atom co-ordinations.

Here the cubic space group Pn3m (224) is the most troublesome. There are 4,500 such entries in the ICSD database, of which ~50% required a correction to their co-ordinates. To improve this situation we have made important modifications to the CDS implementation of ICSD-WWW so it can handle all problem structures properly without manual intervention. We have liaised with Grenoble and hope that these changes will soon be added to the standard version of the ICSD-WWW system released to users worldwide.

Below is a CrystalWeb hit display for one the 19 entries for Cuprite (Cu_2O) in the ICSD database with space group Pn3m (224). Structure determinations date back to 1915, but the example is for a synchrotron study from 1990. This retains the older, now "non-standard" setting, for this space group.



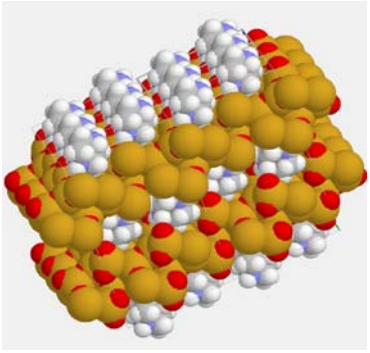
Another recent enhancement shown in the CrystalWeb hits display is that users can make a direct link between hits retrieved from their search and the various useful features provided by ICSD-WWW. This is available for entries from the CSD, CrystMet (metals, alloys and intermetallics), and the ICSD database itself. In the near future we plan to provide a direct link for ICSD-WWW hits to the CrystalWeb system. We will also add an option for users to upload their own co-ordinates sets and thus have access to

(Continued on page 6)

Links between the CrystalWeb & ICSD-WWW database

(Continued from page 5)

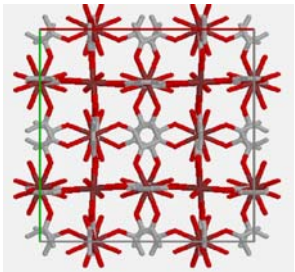
the main features available of both CrystalWeb and ICSD-WWW. Screen shots showing a pyrazinium silicate structure [6] give some indication of display capabilities.



Here CrystalWeb highlights the silicate sheets with the pyrazinium moieties sandwiched between.

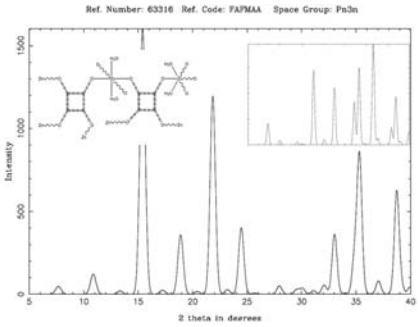
The link to ICSD-WWW allows the silicate part to be displayed using Xtal-3d in polyhedron mode.

Use of the new CrystalWeb/ICSD-WWW features is illustrated by another structure reported in a "non-standard" space group setting. Here we select an entry from the CSD which has the relatively rare space group Pn3n (222). This is an extended zinc "squarate" complex. It was described by the authors as a water/acetic acid clathrate, but the co-ordinates for these molecules were not determined.



A display produced by CrystalWeb gives an indication of the network structure of CSD Refcode entry FAFMAA.

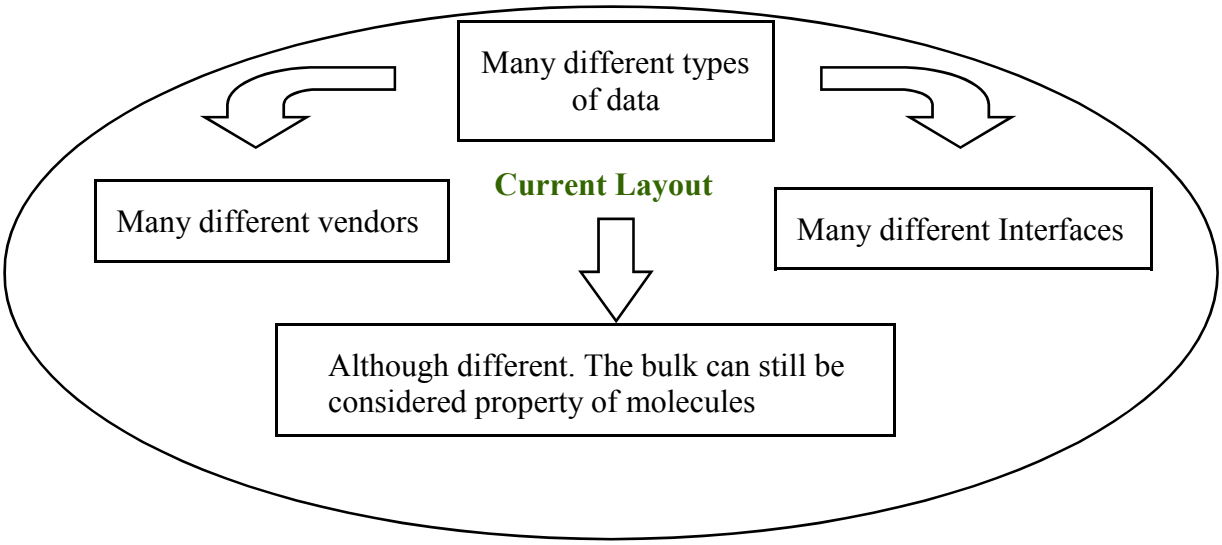
The diagrams show the structural display using CrystalWeb and the simulated neutron and X-ray powder patterns available via the ICSD-WWW link. A later paper [7] recognised the possibility of sodalite networks for certain metal squarates complexes. These new structures are present in the ICSD database and can be found using CrystalWeb. It is then easy to display the structures and powder patterns and demonstrate the zinc compounds from both studies are indeed very similar.



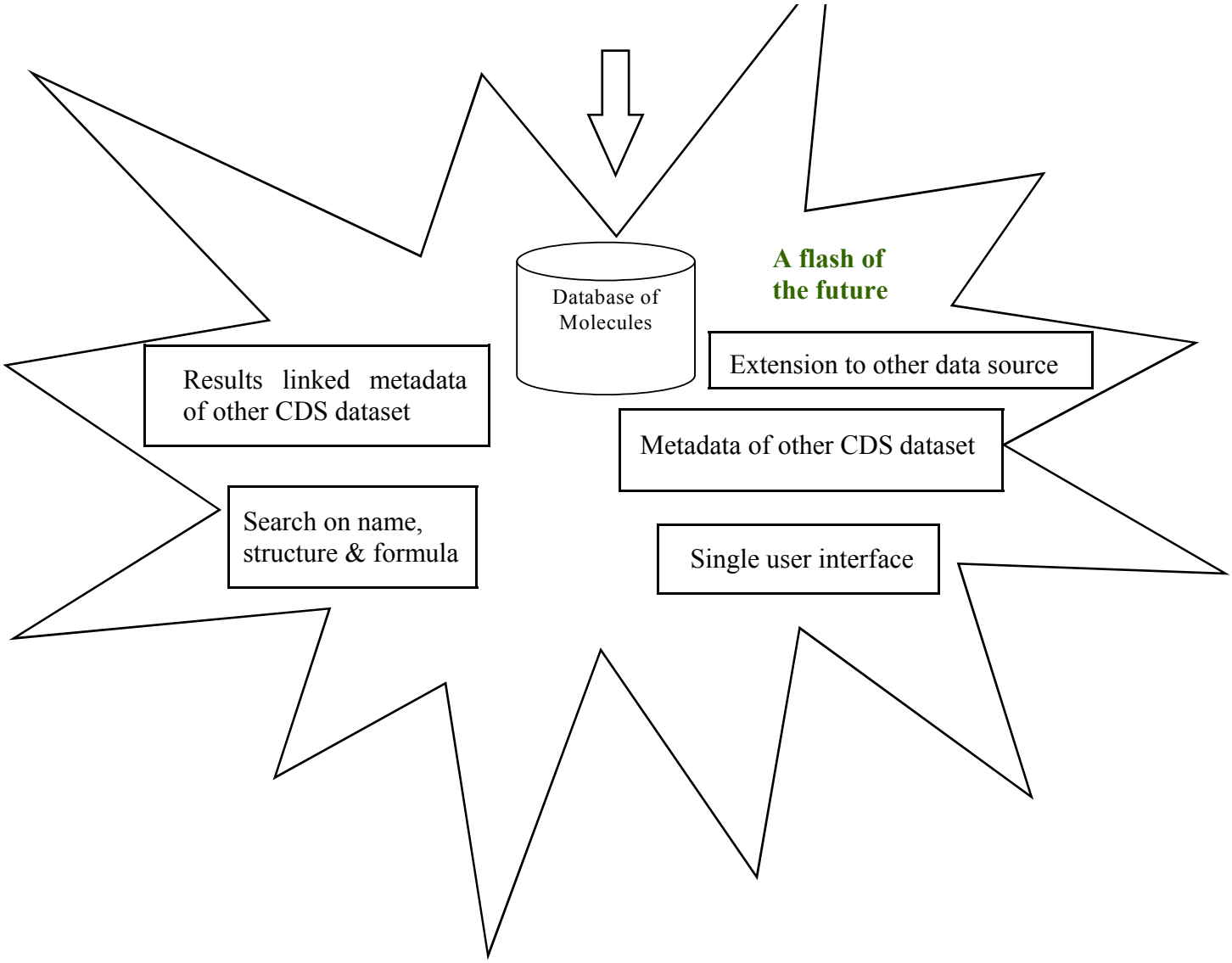
ICSD-WWW was used to generate powder patterns. The main display is for neutrons in the range 2Θ 5-40°. The X-ray pattern at the same wavelength above 25° is in the insert, which is aligned with the main pattern.

[1] "The United Kingdom Chemical Database Service", D.A. Fletcher, R.F. McMeeking, D. Parkin., *J. Chem. Inf. Comput. Sci.* (1996), **36**, 746-749.
[2] "The CrystalWeb database interface", Bob McMeeking & Dave Fletcher, *CDS Research Highlight 2002-3*, http://cds.dl.ac.uk/report/res_high.html;
"The United Kingdom Chemical Database Service: CDS",
B. McMeeking & D. Fletcher, in *Cheminformatics Developments: History, Reviews and Current Research* (Ed. J. H. Noordik), IOS Press, Amsterdam, Chapter 2, pp 37-67, 2004.
[3] An example in archaeology is: "SR-XRD, Culture Heritage and the CDS database" Manolis Pantos, *CDS Research Highlight 2001*, http://cds.dl.ac.uk/report/res_high.html..
[4] "ICSD for WWW", A.W. Hewat, *et al.*, http://icsd.ill.fr/icsd.
[5] http://www.ill.fr/dif/3D-crystals/xtal-3d.html.
[6] "Piperazine silicate (EU 19): the structure of a very small crystal determined with synchrotron radiation", S.J. Andrews M.Z. Papiz R. McMeeking A.J. Blake B.M. Lowe K.R. Franklin J.R. Helliwell M.M. Harding *Acta Cryst.* (1988), B**44**, 73-77.
[7] "Sodalite networks formed by metal squarates", S. Neeraj, M.L. Noy, C.N.R. Rao, A.K. Cheetham, *Solid State Sciences* (2002), **4**, 1231-6.

CDS Molecule Database: The future



Currently in order for CDS users to access different types of data they have to familiarise themselves with various interfaces and keep up to date with improvements to these interfaces. The CDS is currently at the specification/design stage of our Database of Molecules project. This Database of Molecules will have a single and simple user interface. Users will be able to carry out searches on names, structures and formula. Users will have the option of carrying out their search on our entire portfolio of databases or on a subset of our databases.



Acknowledging the Service

The results of research using the Chemical Database Service may be published through the normal academic channels provided the following acknowledgement is quoted:

"We wish to acknowledge the use of the EPSRC's Chemical Database Service at Daresbury."

The following publication should also be cited:

The United Kingdom Chemical Database Service, Fletcher, D.A., McMeeking, R.F., Parkin, D., J. Chem. Inf. Comput. Sci. (1996), **36**, 746-749.

Spectroscopy Data

HNMR Dataset

Name of the new product:
Wiley 1HNMR Spectra of Organic Compounds 2005

Number of records (full spectra): 96,500

(The collections V2 and V3 included about 38,000 records. Both are contained in the new database, almost doubling the records available).

PNMR and FNMR Dataset

PNMR contains 14,378 assigned 31-P NMR spectra and structures

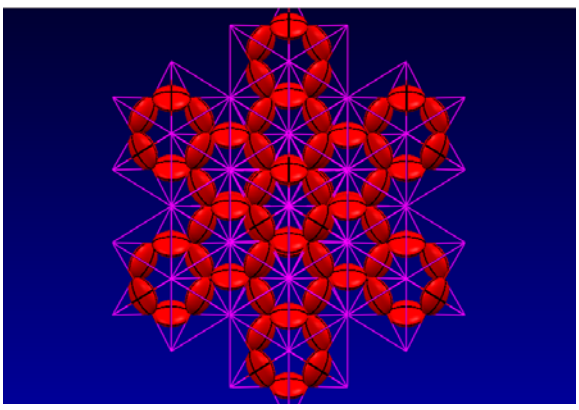
FNMR contain assigned 19-F NMR spectra and structures

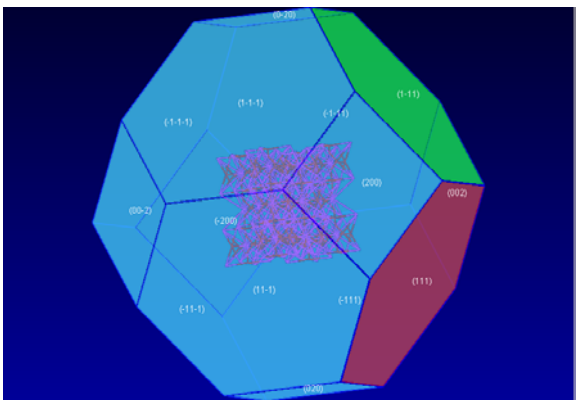
Both datasets increase the number of spectra available **by an order of magnitude**, and should significantly improve the quality of database based P and F NMR predictions.

Cambridge Mercury new features

It is now easy to import structures from the Inorganic Crystal Structure Database

The diagrams below illustrates the thermal ellipsoids





Databases available from the Chemical Database Service (CDS)

A brief description of all the databases currently available from the CDS at Daresbury.

Organic Reactions

They allow you to search, retrieve and display molecules and reactions along with their associated data. Currently there are around 1.4 Million searchable reactions from established, up-to-date and specialist reaction databases.

Available Chemicals

They provide information about compounds that are available for sale and contain information such as supplier, catalogue number, amount of compound, purity and price. The Available Chemicals Directory contains almost 500,000 unique, off-the-shelf compounds from around 680 different suppliers while the Screening Compounds Database contains over 4.2 Million compounds from 24 suppliers.

Structures

They contain crystal structure data on organic, organo-metallic and inorganic compounds as well as metals and alloys. These include:-

CSD :The Cambridge Structural Database - the only comprehensive collection of small-molecule organic and organometallic crystal structures.
ICSD: The Inorganic Crystal Structure DataFile - inorganic structures - the companion file to the Cambridge organic file.
MDF (CRYSTMET) Metals Data File) - Crystal structure data for metals and alloys.
CDIF: Crystal Data Identification File - Crystal class and unit cell data for crystal structures.

Spectroscopy

They are designed to aid the chemist in structure elucidation and spectral interpretation problems. SpecInfo is a database management system designed to store, retrieve, and manipulate NMR, IR, and mass spectra. It comes with an associated high quality dataset comprising 102,000 C-13 NMR spectra, 61,000 proton NMR spectra, 6,000 heteroatom (N-15, O-17, F-19 and P-31) NMR spectra, 130,000 mass spectra and 21,000 IR spectra.

Physical Properties

It provides data on the properties of elements and pure compound as well as mixtures. CDS provide access to **Detherm** which is one of the world's largest thermophysical property databases. It contains over 5 Million data sets for around 118,000 systems (over 24,700 pure substances and 93,500 mixtures) covering around 500 property fields.

Utilities

A variety of utility programs are also made available, including links to electronic literature (**LitLink**), chemical file format conversion (**Babel** and **Bedlam**) and molecule viewers (e.g. **Rasmol; Mercury**).

CDS Helpdesk: cdsbb@dl.ac.uk or 01925-603-162

General Enquiries

Chemical Database Service (CDS)
Daresbury Lab
Daresbury
Warrington
WA4 4AD
(01925) 603162
Fax:(01925) 603031
Cdsbb@dl.ac.uk

CDS website: http://cds.dl.ac.uk

CDS ON-LINE

The Chemical Database Service (CDS) provides on-line access to a variety of quality databases in the field of Chemistry, plus support, training and advice.

Access:
The service is available **free of charge** to UK academic research groups for non-commercial work. Each individual user will be issued with a unique ID. It is not our policy to allow shared IDs.

Remit: To provide Chemical Database Service and support to UK academic community as well as helping to maintain/ improve the service according to the service level agreement with the EPSRC

Information

For ISIS specific problems:

Dr. D. Parkin
Email: D.Parkin@dl.ac.uk
(01925) 603162

For SpecInfo specific problems:

Dr. D.A. Fletcher
Email:D.A.Fletcher@dl.ac.uk
(01925) 603492

For crystallography specific problems:

Dr. R. F. McMeeking
Email:R.F.McMeeking@dl.ac.uk
(01925) 603669

World Wide Web Site

Information about CDS, including online help and documentation is available over the World Wide Web at the CDS website, URL: http://cds.dl.ac.uk Web based interfaces to selected databases are also available from this site.

Documentation:

Most documentation is available online and some can be downloaded from the CDS web site.

cds (main service machine):

Internet name: cds.dl.ac.uk

Comments:

All comments, questions and suggestions about this newsletter should be sent to:
David Osa-Edoh
Email:D.Osa-Edoh@dl.ac.uk
(01925) 603163