

Chemical Database Service (CDS) NEWSLETTER

Issue 20 Oct 2004



See <http://cds.dl.ac.uk/links/news1.html> for this and previous editions of our newsletter

A summary of the main articles

The CDS Road Show 2004/05. See the article on page 2.

For more information on our ongoing trials of Accelry's Failed Reaction, Bioster and Metabolism databases see page 4 and 5.

See page 6, 7 and 8 for news on the latest version of DETHERM.

For the best way to access the non web resources of CDS with a Mac running OSX see Page 9.

Two of our CDS user's research highlights are presented on page 10 and 11

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

Inside this issue:

<i>CDS Road Show</i>	2
<i>Accelrys Data Trials</i>	4
<i>New Detherm V 2.0</i>	6
<i>Mac OSX Access to CDS</i>	9
<i>CDS User's Research Highlights</i>	10
<i>Databases available from the CDS</i>	12
<i>General Enquiries</i>	13

CDS Road Show 2004/05

The Chemical Database Service (CDS) is currently embarking on a tour of universities. The road show will include poster displays, face to face CDS discussion, handing-out of CDS flyers, registrations and a CDS overview presentation. The overview will cover the various chemistry datasets hosted by the service (Physical Chemistry, Organic Chemistry, Crystallography and Spectroscopy) and their access methods (via a Web Based Interface, X-windows or Client-Server).

University	Department	Provisional Date
Oxford	Chemistry	30/09/2004
Warwick	Chemistry	05/10/2004
UCL	Chemistry	06/10/2004
Swansea	Chemistry	13/10/2004
Cardiff	Chemistry	14/10/2004
Nottingham	Chemistry	15/10/2004
York	Biological Sciences	19/10/2004
Leeds	Chemistry	21/10/2004
Sheffield	Chemistry	26/10/2004
Newcastle	Chemical Engineering	28/10/2004
Durham	Chemistry	29/10/2004
Belfast Queen's	Chemical Engineering	04/11/2004
St. Andrews	Chemistry	08/11/2004
Edinburgh	Chemistry	09/11/2004
Dundee	Biological Sciences	10/11/2004
Glasgow	Biological Sciences	11/11/2004
Strathclyde	Chemistry	12/11/2004
Bangor	Chemistry	19/11/2004
Liverpool	Chemistry	25/11/2004
Manchester	Chemistry	30/11/2004
Southampton	Biological Sciences	25/01/2005
Bath	Chemical Engineering	10/02/2005
Bristol	Chemistry	11/02/2005
Loughborough	Chemical Engineering	17/02/2005
Leicester	Chemistry	24/02/2005
Birmingham	Chemistry	01/03/2005
Cambridge	Chemistry	08/03/2005

CDS Road Show

CDS Road Show 2004/05

The overview will cover the various chemistry datasets hosted by the service:

Physical Chemistry

<http://cds.dl.ac.uk/physchem.html>

Organic Chemistry

<http://cds.dl.ac.uk/orgchem.html>

Crystallography

<http://cds.dl.ac.uk/crystal.html>

Spectroscopy

<http://cds.dl.ac.uk/spec.html>

and their access methods (via a web based interface or X-windows).

We will require a seminar room with internet connection (in the afternoon) and a poster stand space with internet connection (in the morning). If you are interested in arranging a course for your department/group, the procedure is to contact your group representative (<http://cds.dl.ac.uk/cgi-bin/reg/findrep.html>) or contact us yourself (if you have no representative). Group representatives or interested/potential users can contact CDS via email: cdsbb.dl.ac.uk or phone: (01925) 603163.

CDS Road Show

Accelrys Data Trials

The CDS carried out trials of Accelrys Failed Reaction, Metabolism and Bioster databases earlier in the summer. The number of responses to the user survey conducted towards the end of the trials was very disappointing, hence we have decided to carry out a re-run of the trials until the end of October 2004. Another user survey will also be carried out towards the end of this trials.

We encourage users to try out these databases and take part in the survey. The result of the survey will be used in deciding whether these databases should be permanently added to our portfolio of hosted databases.

Failed Reactions

Accelrys' Failed Reactions offers a unique resource for the synthetic chemist - the chance to avoid repeating other people's mistakes! The database contains reactions, abstracted from the literature, which did not proceed as expected. These reactions fall into three categories: unexpected product, further reaction, or no reaction where one was expected.

ISIS/Base - [failedrxns.db/Default]

File Edit Options Object Database Search List Window Help

Forms Query Browse Update

<RXN> 8 of 13564
Search Domain: All

Expected Reaction:

Actual Product(s):

Title: Transannular closure of a nine-membered lactam on attempted silyl enol ether formation

Ref.: Desmaele D, Tetrahedron Lett, 37(1), p. 1233, 1996

Text: The ring-closure is an aldol-type condensation induced by the strong base Et₃N. A related nine membered ketolactam underwent similar closure on attempted Wittig reaction of the ketone group with PPh₃-CHCO₂Me.

Reaction Conditions:

Et ₃ N
CH ₂ Cl ₂
20 C

Product: 1 of 2 Yield: %

Experimental Details? No

Record No.: 8 **SDB No.:** RSFR00030008 **Category:** Unexpected reaction

Keyphrases (expected reaction):
Enolisation, O-Protection, O-Silylation

Keyphrases (actual reaction):
Aldol addition, Aldol condensation, O-Protection, O-Silylation, Ring contraction, Transannular

Accelrys Data Trials

The database

Although Failed Reactions can be used independently, it attains its full potential when used in conjunction with other synthetic chemistry databases. This allows you to search for a desired product, then see examples of both successful and unsuccessful experiments. To facilitate this way of working, Accelrys have taken great care in developing the user interface forms to clearly distinguish failed reactions from those in other databases.

All the references are abstracted by highly experienced PhD chemists with experience in seeking out this type of data, so you can be sure of the relevance and quality of the information.

Metabolism

The prospect of any new compound becoming commercially successful, for example as a drug or agrochemical, is critically dependent on its biological activity, including its toxicity and its effects on the environment. An understanding of the likely metabolic fate of a compound can help to predict its activity, thereby influencing its design or prompting design modifications. It is in this context that Accelrys has developed its new Metabolism database.

The database is configured as a **reaction** database in which the starting material is the parent (test) compound and the product a metabolite.

Metabolism provides a must-have reference source for entry into the primary literature on the metabolic fate of organic molecules. It covers biotransformations of organic molecules in a wide variety of species and is compiled in association with experts in the field.

The database may be searched graphically using exact, similar and substructure search options, including changed bond information where appropriate. More generic searches may be performed using a variety of keyword options.

Additionally, the database may be indexed with the Use or Occurrence of the parent compound and the Test System, including the method of administration.

Bioster

Replacement of bioisosterically equivalent moieties is an important and well recognised approach to analog design. A bioisosteric replacement may help optimise activity, remove unwanted side effects, or change a bulk property such as the transportability of the drug or agrochemical.

Bioanalogues are displayed as Reactant and Product, respectively, of a hypothetical 'reaction' (Bioisosteric Transformation), with the analogous, i.e. structurally replaceable, moieties highlighted. It is useful in designing new active molecules, based upon the principle of bioisosterism. The BIOSTER database is a critical compilation of thousands of bio-analogous molecule pairs (including drugs, agrochemicals and enzyme inhibitors), selected from primary literature references. Critical selection of material and compilation of the database is the responsibility of Dr. Istvan Ujvary of the Plant Protection Institute at the Hungarian Academy of Sciences. The database provides keywords indicating the mode of action and cross-referenced information to the papers in which each active compound is reported. This unique compendium is a must for discovery scientists seeking to optimise the biological activity of a lead compound or remove undesirable side effects.

New Detherm (Version 2.0) Coming Soon

DEThERM® is a numerical database containing thermophysical data of pure substances and mixtures. It is funded by DECHEMA e.V. and currently the database contains 4.9 million data sets for about 130,000 systems.

The retrieval software has been completely redeveloped and the database design revamped. The main development aims have been:

- Increased and up-to-date usability
- New retrieval options
- Increased retrieval speed

The CDS is currently working on the technical issues involved with the rollout of this latest version of Detherm to our user community. Users will be notified by email as well as via the CDS website (recent news and physical chemistry sections) when version 2.0 is released.

Main features

The following main features are part of DETHERM V 2.0:

GUI with multiple document interface: One main window containing several views of the data(base).

Scalable GUI: After choosing a font of your choice all GUI items are rescaled. This enhances readability especially for customers using high resolution monitors.

A choice between two search modes: Wizard Search and Extended Search. The Wizard Search is designed for fast data access using the most typical retrieval questions. The Extended Search allows all kinds of complex retrieval questions and is also focused on inventor requests. For example, one will be able to search for components having an azeotrope with substance A and forming a ternary Liquid-Liquid Equilibrium (LLE) with the mixture A/B.

Structured data view: a tree which lists systems and properties is combined with a list of data sets. The list displays all basic information of a data set: content, size (no. of lines), temperature and pressure ranges, author, year of publication. The list can be ordered by any of them.

Joined data view: Data from several sources (data sets) can be displayed together in one table. The table can be sorted by any row or even combination of rows. The values are displayed in user defined units. Unit conversions (from mass to mole base and vice versa) are performed on the fly during data joining. The view is also printable. Larger tables are spread on different pages.

Optimized graphics representation. Drawing of compensating spline curves for direct interpolation without any model.

Multiple threading for database access: Searches can be terminated any time during execution.

Optimization of searches: Instead of using the ORACLE optimizer, DETHERM keeps its own statistical data and uses them for minimizing the response time of queries. For example, if you search for the density of 3-hexanol, the substance is searched first and then the property. If you are looking for the dipole moment of water, the search is done just the other way.

New Detherm (Version 2.0) Coming Soon

DETERM's modern GUI lets you enter complicated inquiries in a very simple way. As an example we present a search for substances that form an azeotrope with dimethyl sulfoxide (DMSO, 1,1'-sulfinylbismethane) and also have a LLE with water. This complicated inquiry can be specified in just a few steps:

1. Opening the "Extended Search" mask.
2. Selecting the search for a pure substance
3. Selecting the tab for specifying a property
4. Specifying the property "azeotrope"
5. Now specify the substance, the searched substance should form an azeotrope with.
6. The "use" button adds the specified property "azeotrope with sulfinylbismethane (DMSO)" to the query and clears the property page for further input.

The user interface

The screenshot displays the Detherm GUI with several callouts indicating search steps:

- 1. Opening the „Extended Search“ mask**: Points to the 'Extended search' window title.
- 2. Search for a pure substance**: Points to the 'Pure substances' radio button in the 'Search for' section.
- 3. Open the properties page to specify properties**: Points to the 'Properties' tab in the 'Search term' section.
- 4. Substance has an azeotrope**: Points to the 'azeotrope' text in the search term input field.
- 5. After this button has pressed, the box for specifying the second substance, with which the search substance has an azeotrope, is opened**: Points to the 'Choice of related component(s)' dialog box.

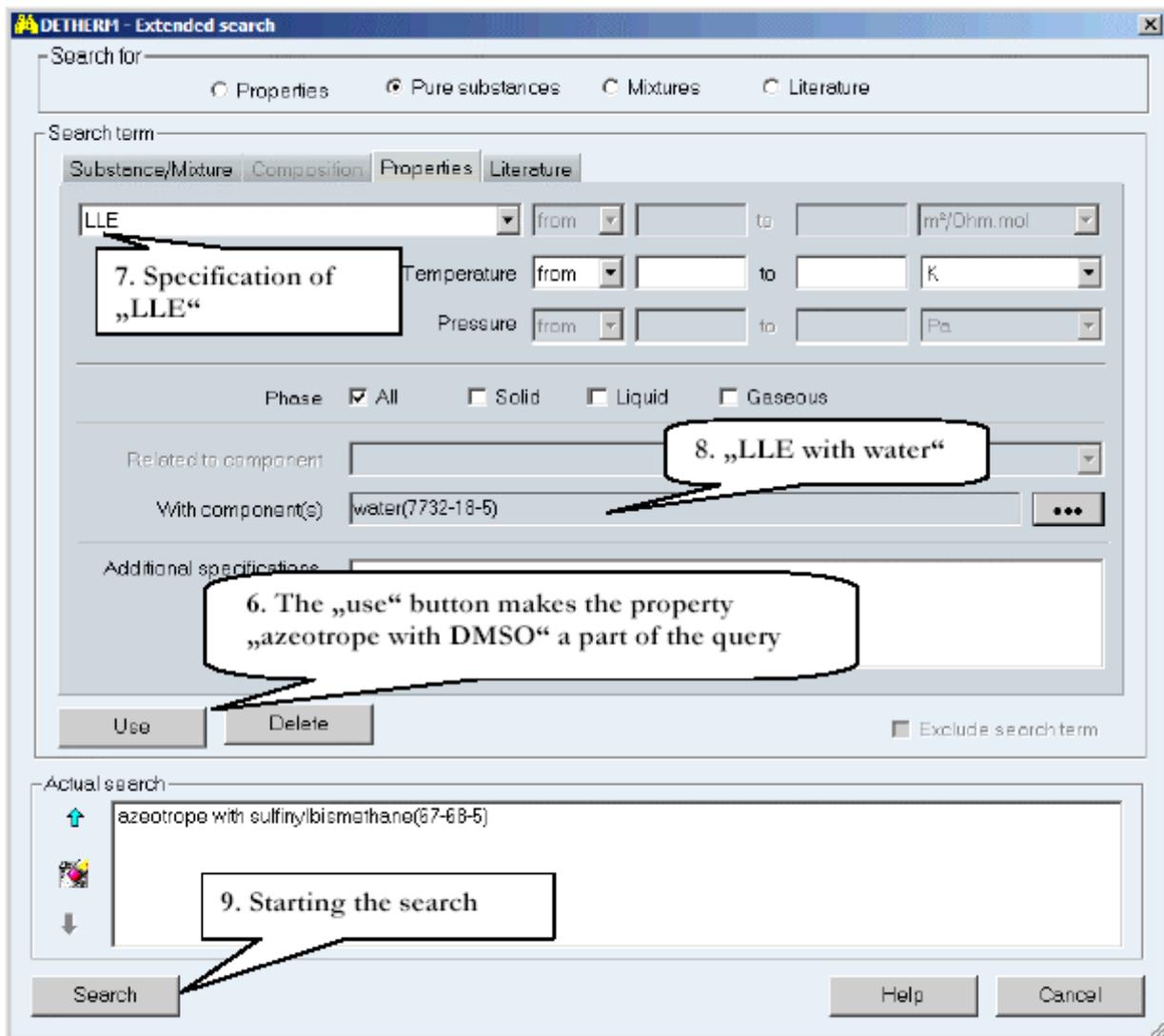
The 'Choice of related component(s)' dialog box shows a list of substances with the following columns: Hit, Systemic name, and Sum. The search term 'dmsa' is entered in the input field.

Hit	Systemic name	Sum
DMSO	sulfinylbismethane	C2H6S
DMSO-D6	sulfinylbismethane-D6	C2D6S
DMTP (insecticides)	phosphorodithioic acid S-(5-methoxy...	C8H11PS2
docosfluorodecane	docosfluorodecane	C10F22
docosmethylcycloundecasiloxane	docosmethylcycloundecasiloxane	C22H44OSi
docosmethyldecylloxane	1,1,1,3,5,7,9,9-nonaethyl-5-[(1,3,3,3...	C22H44OSi
docosone	docosone	C22H44
docosenic acid	docosenic acid	C22H44O2
docosenic acid 2-hydroxyethyl ester	docosenic acid 2-hydroxyethyl ester	C24H48O3
docosenic acid ethyl ester	docosenic acid ethyl ester	C24H48O2
Docosenic acid labeled with carbon...	Docosenic acid labeled with carbon...	D172606830
docosenic acid methyl ester	docosenic acid methyl ester	C23H46O2
docosenol-(1)	1-docosenol	C22H46O
Docosic acid	docosenic acid	C22H44O2
docosyl bromide	1-bromodocosane	C22H45Br
docosylbenzene	docosylbenzene	C28H50

Mixture(s) with 1 to components, existing of...

New Detherm (Version 2.0) Coming Soon

- The second property, the searched substance should have, is “LLE (liquid-liquid equilibrium)”
- The partner for the LLE is water. This is specified as shown above for DMSO.
- The search is started by pressing the “search” button. Only two substances match the request: *5-nonanone* and *acetic acid*.
- To display the data of one of the substances simply select a substance and press the “property” button to open the data view.



Mac OSX Access to CDS

The best way to access the non web resources of CDS with a Mac running OSX is to use the X11 package from Apple. It comes as part of the OS 10.3 or higher operating system and it can also be freely downloaded from Apple (<http://www.apple.com/downloads/macosx/>).

Once started, X11 should produce a terminal window, if not select Terminal from the Applications menu. Click in this window and type the following:

```
ssh -X cxxx@cds.dl.ac.uk
```

where cxxx is replaced by your CDS id.

You should then be prompted for your CDS password. Once you have logged in you will see the cds% prompt. Type "xlogo" to test the X-Windows connection. If it is working correctly you will see a small window appear with an X in it.

Increasingly, site firewalls are blocking incoming unencrypted X-Windows traffic for security reasons. The above method of access avoids this problem, since the X-Windows traffic is sent encrypted through the ssh connection. However, it is important, once you have logged in, that you DO NOT set the DISPLAY environment variable with setenv as you would normally to use X-Windows. If you do, the X-Windows traffic will be sent direct and unencrypted and may be blocked by your site firewall.

CDS Users Research Highlights

The CDS implementation of the DETHERM Thermophysical Database in the Imperial Chemical Engineering Department – Amparo Galindo (a.galindo@ic.ac.uk)

Advanced Research Fellow of the EPSRC

I work as a Lecturer in the Department of Chemical Engineering and Chemical Technology in Imperial College London, and I am also an Advanced Research Fellow of the EPSRC. The research interests of my group focus on the development of statistical mechanical approaches for complex systems, and their application to processes relevant to industry. Our interest in this field is to develop fundamental approaches to contribute to the understanding of experimental systems, with a special focus on chemical processes. The goal is to be able to truly predict complex phase behaviour. The types of problems we are interested in at the moment include charged systems, near-critical and supercritical separations, mixtures of polymers and liquid crystals, and solid phases of chain molecules. The impact and exposure of this work is maximised through collaborative efforts in which the aim is to promote the transfer of the theoretical developments into tools for the design and synthesis of chemical processes and products.

In carrying out this work we rely ultimately on experimental information of thermodynamic properties of the systems of interest: sometimes to correlate the parameters in our models, and sometimes to assess the accuracy of our theoretical predictions. Although we collaborate within and outside the UK with groups involved in experimental work, in most cases the data we use is from previous published work. The search for accurate and relevant data can become an intensive task, and in this sense the Chemical Database has become an invaluable tool in our group. It saves incredible amounts of time to have on-line access to such a compilation of information. Two especially useful characteristics of the database are that it provides full references to the sources of the data for verification and assessment of experimental techniques, and that the data can be tabulated or formatted into graphic form and transported electronically directly into Windows or Linux packages.

Patel BH, Paricaud P, Galindo A, Maitland GC (2003) "Prediction of the salting-out effect of strong electrolytes on water plus alkane solutions", *Ind. Eng. Chem. Res.*, **42**, 3809-3823

X-ray diffraction studies of hydrous carbonates - Alistair Lennie (a.lennie@dl.ac.uk), Chiu Tang (c.c.tang@dl.ac.uk) & Steve Thompson (s.p.thompson@dl.ac.uk)

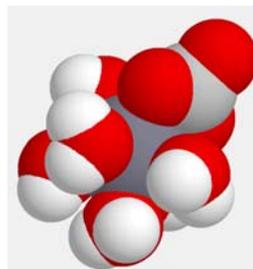
Work supported by the EPSRC

Changes in carbon dioxide in the atmosphere over geological time are interconnected with organic and mineral carbon reservoirs. The predominant carbonate mineral calcite (CaCO_3) acts as a buffer for long-term cycling of carbon dioxide between the atmosphere, oceans and solid Earth. However, calcite is not the only calcium carbonate mineral, and there are still discoveries to be made in this complex and fascinating mineral system. In 1963, a then little-known calcium carbonate phase was reported as a new mineral, ikaite, in unusual underwater formations at Ikka Fjord, Greenland. Although ikaite was first discovered as a mineral in 1963, it had been reported from laboratory studies made during the 19th century. Ikaite plays a role in the carbon cycle, sequestering carbon dioxide produced by bacterial oxidation of organic matter in cold marine sediments. How important this role is in the total carbon dioxide budget of the Earth's oceans has not yet been established.

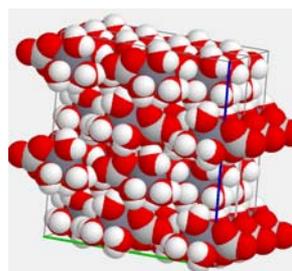
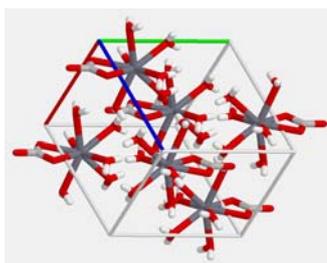
Ikaite, although metastable at low pressures, will form below about 278 K if there is sufficient dissolved calcium and carbonate available. This explains its presence in sediments in Bransfield Strait, Antarctica, and in the Zaire deep sea fan, as well as its occurrence at Ikka Fjord. This mineral thus provides a sink for carbon dioxide in Arctic, Antarctic and deep-sea marine sediments. Using synchrotron radiation, we are measuring properties of this mineral at low temperature and at high

CDS Users Research Highlights

pressure, and are following processes of decomposition as temperature is increased. Our studies investigating the behaviour of this unusual mineral are helping us to understand the role of ikaite in the Earth's CO₂ cycle. The crystal structure of ikaite consists of repeating units of six water molecules, a carbonate anion and a calcium ion. Oxygens from six water molecules, and two oxygens from carbonate, coordinate calcium to form calcium carbonate hexahydrate units (CaCO₃·6H₂O). This description of the structure becomes clearer when viewing the structure in three dimensions, possible with the aid of CrystalWeb, the Chemical Database on-line facility for crystallographic display



In ikaite, hydrogen atoms belonging to water molecules link with nearby oxygens to form hydrogen bonds. These weak bonds join individual units together to form the crystal, similar to the formation of ice which consists of a network of hydrogen-bonded water molecules.



We synthesised crystals of CaCO₃·6H₂O and examined these at the SRS, Daresbury. The data collected has shown us how, as the crystal cools, weak hydrogen bonds in the structure undergo greatest change. When we compare ikaite with ice (H₂O) and gypsum (CaSO₄·2H₂O), we observe that ikaite thermal behaviour lies in-between these minerals. Here again, the databases of CDS provide invaluable reference data from which we can build up a picture of comparative behaviour. Gypsum has significantly less hydrogen bonding within its structure than ikaite, while ice has more, with water molecules linked by hydrogen bonds only. These structural differences are reflected in their crystallographic behaviour. When ikaite is warmed to room temperature, it breaks down into calcite (CaCO₃) and vaterite (CaCO₃), another rare calcium carbonate phase, with water of crystallisation being released. We are now following the formation of these new phases using X-rays to track down structural features in common, and to understand mechanisms of formation and transformation. Again, the databases are of great help here, allowing easy visualisation of crystal structural relationships. An unusual feature of this icy carbonate is that the molar volume is 21% smaller than that occupied by calcite + 6 H₂O liquid phase. Under high water pressure therefore, ikaite becomes a stable phase, and we are using high-pressure studies to find out how this structure is stabilised. Again, we can compare this with other high-pressure phases recorded in ICSD.

Hopefully in the future we will be able to contribute to the database of structures from which we are at present drawing by providing new structural information to add to the present database.

AR Lennie, CC Tang and SP Thompson (2004) "The structure and thermal expansion behaviour of ikaite, CaCO₃·6H₂O, from T = 114 to T = 293 K". *Mineralogical Magazine*, **68**, 135–146.

Databases available from the Chemical Database Service

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

ISIS: A chemical reaction information management system allowing search, retrieval and display of molecules, reactions and their associated data.

Currently contains over 1,300,000 searchable reactions. Specialist Databases available include Protecting Groups, Solid Phase Synthesis, BioCatalysis, Chiral Separations and National Cancer Institute Database. This

database is produced using data and structures from the National Cancer Institute (NCI) Developmental Therapeutic Program (DTP).

In addition there is the Available Chemicals Directory (**ACD**), which contains supplier information on over 300,000 different compounds from over 660 different suppliers.

Screening Compounds Catalogues

Made up of over 3 million compounds available from around 20 suppliers. Such as **SALOR** (Sigma-Aldrich Library of Rare Chemicals), **MAYBRIDGE** and **BIONET**.

SPEC: SpecInfo is a multi-technique spectroscopic database package intended to cover NMR, IR and mass spectra. The package is designed to aid the chemist in spectral interpretation and structure elucidation problems.

The database currently contains 102,500 ¹³C NMR; 999 ¹⁵N NMR; 856 ¹⁷O NMR; 2,183 ³¹P NMR; 1,825 ¹⁹F NMR; 60,700 ¹H NMR; 20,898 infra-red spectra and 130,000 mass spectra.

CSD: The Cambridge Structural Database. Crystal structure data for over 322,000 organic and organometallic compounds. Currently available under Cambridge's Quest and ConQuest retrieval software and our in-house software (CSSR).

VISTA is a **VI**sual **STA**tistics package which reads the tables files which are when derived geometric parameters are requested with a CSD search.

Mercury offers comprehensive facilities for visualising crystal structures in three dimensions. Both VISTA and Mercury can also be called from within

ConQuest.

IsoStar: A database of non-bonded interactions taken from the CSD, PDB and theoretical calculations.

ICSD: Inorganic Crystal Structure Data File. Contains about 76,000 inorganic structures - the companion file to the Cambridge organic file. ICSD is now accessible via CrystalWeb on the CDS website.

Mogul: A knowledge Base of molecular Geometry using data derived from the CDS

MDF: (CRYSTMET) Metals Data File. Crystal structure data for nearly 75,000 metals, alloys and intermetallics.

CDIF: Crystal Data Identification File. Crystal class and unit cell data for over 237,000 crystal structures.

DETERM: Is one of the world's largest thermophysical property databases of pure compounds and compound mixtures.

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.

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 number: 193.62.124.35
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