

# Chemical Database Service (CDS) NEWSLETTER

Issue 19 Apr 2004



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

## A summary of the main articles

CDS grant renewal, request for Research Highlights and acknowledging CDS (Page 2).

See page 3 for news on the return of DETHERM.

CDS Road Show starts in April 2004 See the article on page 4.

For more information on our planned data trials of Accelry's Bioster and Metabolism Database as well as 31PNMR and 19FNMR data from Chemical Concepts, see page 5,6 and 7.

CDS users can now change their password and update their details via a html form. (Page 8).



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

## Inside this issue:

CDS grant renewal	2
Detherm	3
CDS Road Show	4
New Data Trials	5
New Data Trials	6
New Data Trials	7
Hardware Upgrade	8
Databases available from the CDS	9
General Enquiries	10

## CDS Databases

**ISIS:** A chemical reaction information management system.

**SPEC:** SpecInfo is a multi-technique spectroscopic database

**ICSD:** Inorganic Crystal Structure Data File.

**MDF:**(CRYSTMET) thermophysical property Metals Data File. It databases consists of Crystal structure data.

**CDIF:** Crystal Data Identification File. **CSD:** The Cambridge Structural Database. Consisting of Crystal class and unit cell data.

**ELYS:** Electrolyte Solutions Database. **IsoStar:** A database of non-bonded interactions.

**Detherm:** Is one of the world's largest

## Special points of interest:

- CDS Grant Renewal
- The Return of Detherm
- CDS Roadshow
- New Data Trials
- Hardware Upgrade

## CDS Grant Renewal

CDS is please to announce that the EPSRC has decided to renew our grant. This secures the short-term future of the service. I would like to take this opportunity to thank all our users who provided us with research highlights (which are a prerequisite of our grant application) and those that took part in our survey to ascertain the usefulness of the service to the UK's academic community.

## Research Highlights

We would like our users to send in research highlights of no more than 600 words (word or pdf document), ideally written for a non-specialist but scientifically literate audience. The format is similar to that of the research summary requested from EPSRC-funded researchers in their final reports. Our current collection of research highlights as used in our annual reports can be viewed at [http://cds.dl.ac.uk/report/res\\_high.htm](http://cds.dl.ac.uk/report/res_high.htm).

## 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.

***CDS Grant Renewed by EPSRC***

# Detherm: The Return of the king of Physical Chemistry Databases

As a result of a successful trial and favourable feedback from our Detherm user survey, CDS applied to the EPSRC for long term funding of Detherm. The funding application was strongly supported by the referees, and went forward to the EPSRC prioritisation panel. Unfortunately it failed at this stage in the face of strong competition for the limited funds available in the funding round. Accordingly the Detherm service was terminated in November 2002. After re-assessing the situation, Detherm's funding application was incorporated into our service funding renewal application. The successful renewal of our grant means we can resume our hosting of the database. The re-launch date of Detherm is the 1st of April 2004.

<b>Phase Equilibrium Data</b> (e.g. Vapour-Liquid-Equilibria, Liquid-Liquid-Equilibria, Solid-Liquid-Equilibria, Activity Coefficients at Infinite Dilution, Gas Solubilities, and Vapour Pressures)	<b>Transport Properties</b> (e.g. Viscosities, Thermal Conductivities, and Diffusion Coefficients)
<b>Excess Properties</b> (e.g. Excess Enthalpies, Excess Heat Capacities)	<b>PVT-Data</b> (e.g. Densities, Compressibility, Critical Data)
<b>Caloric Properties</b> (e.g. Specific Heat Capacity, Enthalpy, Entropy)	<b>Electrolyte Data</b> (e.g. Conductivities, Transference Numbers)
<b>Interfacial Properties</b> (e.g. Surface Tension)	<b>Azeotropic Data</b>

Total: **211** different Properties

## Contents of a Dataset

### Identification of substance(s)

- Systematical name, synonyms, sum formula, CAS-No., descriptors.

### Numerical data

- Data, description of data, errors, remarks.

### Literature sources

- Author, Journal, Volume, Year, Page, Coden, abstract.

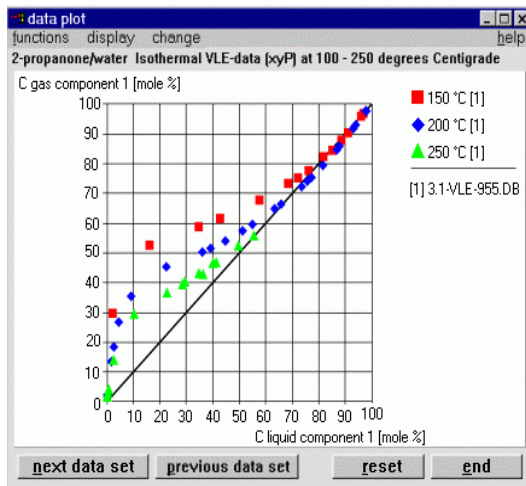
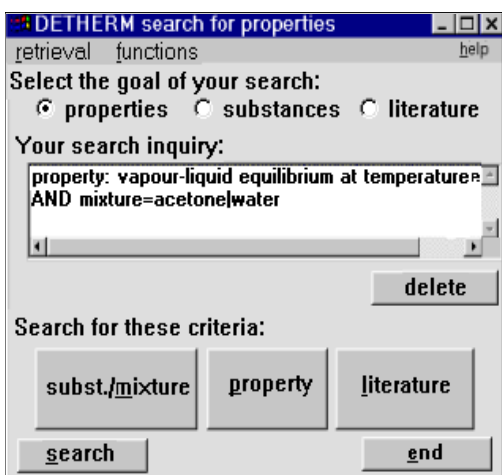
## Technical aspects

The Detherm database runs under a client/server system, similar to the ISIS software. The data is stored inside an Oracle database running on one of the CDS servers, whilst the client software used to access the data runs on your local PC.

You will be able to download the software, free of charge, from the CDS website in a

similar fashion to the ISIS client. The Detherm client software only runs under Windows, and this situation is not likely to change.

More information about the database is available on the Dechema website at the following URL: <http://www.dechema.de>



*The re-launch of  
Detherm*

*1st of April 2004*

## CDS Road Show 2004/05

Starting in April 2004 the Chemical Database Service (CDS) will be 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/Inorganic Chemistry, Crystallography and Spectroscopy) 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 or contact us yourself (if you have no representative). Group representatives or interested/potential users can contact CDS via email: [cdsbb.dl.ac.uk](mailto:cdsbb.dl.ac.uk) or phone: 01925603163.

<b>University</b>	<b>Department</b>	<b>Provisional Date</b>
Sussex	Chemistry	20/04/2004
Imperial	Chemical Engineering	21/04/2004
Warwick	Chemistry	05/10/2004
UCL	Chemistry	06/10/2004
Swansea	Chemistry	12/10/2004
Cardiff	Chemistry	13/10/2004
Nottingham	Chemistry	14/10/2004
York	Biological Sciences	19/10/2004
Leeds	Chemistry	21/10/2004
Sheffield	Chemistry	26/10/2004
Newcastle	Chemical Engineering	28/10/2004
Bath	Chemical Engineering	30/10/2004
Durham	Chemistry	02/11/2004
Belfast Queen's	Chemical Engineering	04/11/2004
St. Andrews	Chemistry	09/11/2004
Glasgow	Biological Sciences	11/11/2004
Edinburgh	Chemistry	16/11/2004
Strathclyde	Chemistry	17/11/2004
Dundee	Biological Sciences	18/11/2004
Liverpool	Chemistry	25/11/2004
Manchester	Chemistry	30/11/2004
Southampton	Biological Sciences	25/01/2005
Bristol	Chemistry	27/01/2005
Loughborough	Chemical Engineering	17/02/2005
Leicester	Chemistry	24/02/2005
Birmingham	Chemistry	01/03/2005

*CDS Road Show, starts in April 2004*

## New Data Trials

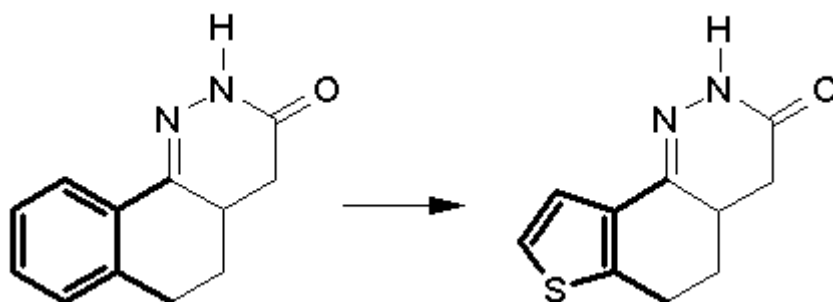
This year CDS plans to run trials of Accelrys' Biooster and Metabolism Databases as well as  $^{31}\text{P}$  NMR and  $^{19}\text{F}$  NMR (Spec-Data) from Chemical Concepts.

### Accelrys' Biooster Database

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

The BIOSTER database, from Accelrys, is a critical compilation of several thousand bio-analogous molecule pairs (including drugs, agrochemicals and enzyme inhibitors), selected from several thousand references published mainly over the last 35 years. Critical selection of material and compilation of the database is the responsibility of Dr. István Ujváry 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.

Although not a 'true' reaction database, bioanalogues are displayed as Reactant and Product, respectively, of a hypothetical 'reaction' (Bioisosteric Transformation), with the analogous, i.e. **structurally replaceable, moieties highlighted**. These are indicated in bold in the following example where a 1,2-disubstituted benzene ring is replaced by a 1,2-disubstituted thiophen:



Accelrys has extracted the part-structures defined by the 'changed bonds' and has indexed them as associated 'fragments'. This greatly enhances the structure searching capabilities within the ISIS system, compared with previous versions of the database

### Accelrys' Metabolism Database

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 Metabolism database.

The advantages of using a database over conventional information sources clearly reside in its search capabilities. Users are able to search not only by generic keywords but also using powerful structure, substructure and similarity features.

## New Data Trials

### Accelrys' Metabolism Database

#### Data Quality

Metabolism data is relevant to a wide variety of disciplines and organisations, including chemists, biologists, toxicologists and various regulatory authorities and government bodies. Often, non-chemists, whose focus has little bearing on the actual structures of the metabolites themselves, publish results of metabolic studies. In creating the database, therefore, Accelrys decided to rely on expert scientists, both dedicated and familiar with abstracting accurate chemical information from a wide variety of primary literature sources.

Accelrys' Metabolism database is derived from the popular book series, Biotransformations, published by the Royal Society of Chemistry. The series continues to be compiled by scientists at HLS, with significant contributions from Zeneca, Pfizer, GlaxoWellcome, AgrEvo and the Medical Research Council. The work is compiled under the expert guidance of Dr. David Hawkins at Huntingdon Life Sciences in the UK. It is now in its seventh volume and future releases of the database will be derived from subsequent volumes of Biotransformations.

Accelrys also plans to augment the database with material from other sources, yet to be defined.

Use of the database as a knowledge base with toxicology prediction software is also under consideration.

#### The Database

The database comprises biotransformations of chemical entities, including pharmaceuticals, agrochemicals, food additives, and environmental and industrial chemicals in vertebrates within the animal kingdom (including mammals, birds and fish). Future releases are scheduled to significantly expand the scope of the database, culminating in a comprehensive survey of the metabolism literature.

The Metabolism database is organised to show transformations of the parent molecule to each individual metabolite. Actual or hypothetical metabolic pathways are indexed alpha-numerically. Additionally, the database is indexed with the Use or Occurrence of the parent compound and the Test System, including the method of administration. 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.

It is anticipated that the database will prove useful to chemists, biochemists and toxicologists and to regulatory authorities and government bodies.

#### KEY FEATURES

- Developed in association with the RSC and expert scientific collaborators.
- Includes drugs, agrochemicals, food additives and industrial and environmental chemicals.
- Searchable by exact and similar structure, substructure, text and keywords.
- Backlog data and regular updates scheduled.
- 

#### Organisation of the Database

The metabolic transformations are presented as a series of chemical reactions whereby the reactant is always the parent compound and the product a metabolite.

## New Data Trials

### <sup>31</sup>P NMR Collection

The <sup>31</sup>P NMR spectral collection contains 14,378 <sup>31</sup>P NMR spectra and chemical structures. Most of the chemical shifts are explicitly assigned to the corresponding phosphorus atoms in the structure. This is the basis of a <sup>31</sup>P NMR spectrum prediction tool.

In addition to structural formula and spectrum, the following information is available:

- Chemical name
- Molecular formula
- Molecular weight
- Literature reference
- Solvent
- Standard
- Measurement condition

### <sup>19</sup>F NMR Collection

The <sup>19</sup>F NMR spectral collection contains 23,617 <sup>19</sup>F NMR spectra and chemical structures. Most of the chemical shifts are explicitly assigned to the corresponding phosphorus atoms in the structure. This is the basis of a <sup>19</sup>F NMR spectrum prediction tool.

In addition to structural formula and spectrum, the following information is available:

- Chemical name
- Molecular formula
- Molecular weight
- Literature reference
- Solvent
- Measurement condition
- Standard

# Hardware Upgrade

## Password Change Form

CDS users can now change their password via a html form. At our website (cds.dl.ac.uk ) the form can be located by placing your mouse over "Registration/Password" menu item on the left-hand site navigational menu and clicking-on "Change your password" on the pop-up sub-menu. Users can now also update their details.



## CDS password change form

ID

Old password

New password

New password (again)

Passwords must be 6-8 characters long and can consist of letters, numbers and the following non-alphanumeric characters.

`_!#$%&()*+,-./:;<=>?@[]^`{|`

Each password must contain at least one non-alphanumeric character as well as both upper and lower case characters.

### Don't give or share your password

Don't let other users use your ID and password. Don't let anyone watch you enter your password. Use the password for a limited time and change it periodically.

# 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 1,200,000 searchable reactions. Specialist Databases available include Protecting Groups, Solid Phase Synthesis, BioCatalysis and Chiral Separations.

**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 99,059 <sup>13</sup>C NMR; 999 <sup>15</sup>N NMR; 856 <sup>17</sup>O NMR; 2,183 <sup>31</sup>P NMR; 1,825 <sup>19</sup>F NMR; 26,000 <sup>1</sup>H NMR; 20,898 infra-red spectra and 128,000 mass spectra.

**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.

**ELYS:** Electrolyte Solutions Database. Thermodynamic and transport property data such as density, viscosity and diffusion coefficients. Currently contains about 11,000 entries.

In addition there is the Available Chemicals Directory (ACD), which contains supplier information on over 254,383 different compounds from over 656 different suppliers and National Cancer Institute Database. This database is produced using data and structures from the National Cancer Institute (NCI) Developmental Therapeutic Program (DTP).

**CSD:** The Cambridge Structural Database. Crystal structure data for over 298,000 organic and organometallic compounds.

Currently available under Cambridge's Quest and ConQuest retrieval software and our in-house software (CSSR).

VISTA is a VISual STATistics 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.

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

## Screening Compounds Catalogues

Made up of three catalogues (removed from ACD). They are **SALOR** (Sigma-Aldrich Library of Rare Chemicals), **MAYBRIDGE** and **BIONET**.

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 67,200 inorganic structures - the companion file to the Cambridge organic file. ICSD is now accessible via CrystalWeb on the CDS website.

## 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  
01925603162  
Fax:01925603031  
Cdsbb@dl.ac.uk

**CDS website: [http://  
cds.dl.ac.uk/cds](http://cds.dl.ac.uk/cds)**

**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](mailto:D.Parkin@dl.ac.uk)  
(01925)603162

**For SpecInfo specific problems:**

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

**For crystallography specific problems:**

Dr. R. F. McMeeking  
Email: [R.F.McMeeking@dl.ac.uk](mailto: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/cds>  
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. Printed manuals can still be ordered from us at the above address.

**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](mailto:D.Osa-Edoh@dl.ac.uk)  
(01925)603163