

Chemical Database Service (CDS) NEWSLETTER

Issue 23. Apr 2006



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

A summary of the main articles

CDS forthcoming trials of SPRESI, ACD/ILAB and Science of Synthesis databases. (page 2-4)

1.3 and 1.4 are advised to upgrade to the latest version. The switch-off date is 18/4/06. See page 6 for more.

The CDS Road Show and training programme is presented on page 5.

Summary of the 2005 user survey (page 7)

See page 6 for CDS user forum 2006.

For how to acknowledge the service (page 7)

The latest version of Detherm (2.1) is now available. Users using detherm client version



EPSRC

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

Inside this issue:

CDS database trials	2-4
CDS Road show & training programme	5
CDS User Forum	6
Detherm upgrade	6
2005 User Survey Feedback	7
Acknowledging CDS	7
Available Databases	8
General Enquiries	9



Database Trials: SPRESI, ACD/ILab and Science of Synthesis.

Users will be notified about access methods by email and on the news section of our website

SPRESI : Trial commencing the 24th of April '05 and ending the 24th of May

SPRESI stands for SPEicherung und REcherche Strukturchemischer Information (Storage and Retrieval of Chemical Structure Information), because historically SPRESI was built as a structural database.

SPRESI is a comprehensive database consisting of chemical structures and reaction information gathered since 1974 from over 600,000 references and 164,000 patents. It providing direct access to over 5.0 million structures, 3.7 million reactions and also includes 28 million factual data, which consist of physical properties (boiling point, melting point, decomposition, density, etc), reaction conditions (catalysts, solvents, etc) and keywords.

Overview

In the 70's, the All-Union Institute of Scientific and Technical Information of the Academy of Science of the USSR (VINITI) in Moscow, and the Information Processing Centre for Chemistry (ZIC) in Berlin started to abstract structures and reactions from the most important organic synthesis journals, creating SPRESI. Since then, abstraction of the world's chemical literature continues to maintain this database up to date.

InfoChem GmbH holds the rights on this data and, since 1990, has made the data available to the west.

What kind of data does SPRESI contain?

In addition to the structures and reactions, the SPRESI Property File provides nine numerical, physical properties and a set of keywords. These data are:

Boiling point

Density

Dissociation constants

Melting point

Decomposition point

Optical rotation

Polymorphic transition point

Refraction index

Sublimation point

Keywords.

ACD/ILab : Trial commencing April '05 and ending July '05

ACD/ILab (ACD Interactive Lab) is a universal Web-based gateway to spectroscopy information, compound name generation and property prediction programs.

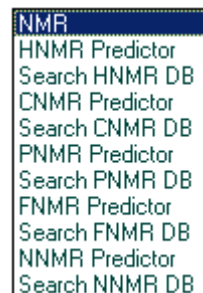
Available through ILab:-

NMR

- ^1H NMR spectrum prediction
- HNMR database search (over 165,000 structures)
- ^{13}C NMR spectrum prediction
- CNMR database search (over 165,000 structures)

Database Trials : ACD/ILab

- ^{19}F NMR spectrum prediction
- FNMR database search (over 13,800 structures)
- ^{31}P NMR spectrum prediction
- PNMR database search (over 22,600 structures)
- ^{15}N NMR spectrum prediction
- NNMR database search (over 8000 structures)



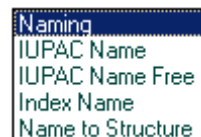
NMR
HNMR Predictor
Search HNMR DB
CNMR Predictor
Search CNMR DB
PNMR Predictor
Search PNMR DB
FNMR Predictor
Search FNMR DB
NNMR Predictor
Search NNMR DB

Naming

- Name generation according to IUPAC rules
- Index name generation according to CAS rules
- Structure generation from chemical name

[Capabilities now include the ability to generate IUPAC and Index (CAS) chemical names for:

- Structure- and source-based names for polymer structures
- Coordination compounds including:
 - Neutral and anionic compounds
 - Cyclopentadienyl complexes like ferrocenes
- Expanded naming of natural and biochemical compounds to include:
 - Porphyrins, phthalocyanines, porphyranzines, and phorbines
 - Nucleotides and nucleosides]



Naming
IUPAC Name
IUPAC Name Free
Index Name
Name to Structure

Molecular Physical Properties (PhysChem)

- pK_a prediction
- pK_a database search (about 16,000 structures)
- $\text{Log}P$ prediction
- $\text{Log}P$ database search (over 18,000 structures)
- $\text{Log}D$ prediction
- Aqueous solubility prediction

Database Trial : ACD/ILab

- Solubility database search (over 5000 compounds)
- Boiling point/Vapour pressure/Enthalpy of vaporization prediction
- Adsorption coefficient/Bioconcentration factor prediction
- Molar Refractivity prediction
- Molar Volume prediction
- Parachor prediction
- Refractive Index prediction
- Surface Tension prediction
- Density prediction
- Dielectric Constant calculation
- Polarizability prediction
- Monoisotopic Mass prediction
- Nominal Mass prediction
- Average Mass prediction

Phys-Chem
pKa
Search pKa DB
LogP
Search LogP DB
LogD
Search Sol. DB
Aqueous Solubility
Boil.Point/Vap.Press
Molar Refractivity
Molar Volume
Parachor
Refractive Index
Surface Tension
Density
Dielectric Constant
Polarizability
Monoisotopic Mass
Nominal Mass
Average Mass
Ads.Coeff./BCF

Science of Synthesis: Trial Access coordinated by the CDS

We have recently been contacted by Georg Thieme Verlag KG, who produces the reference work *Science of Synthesis* (Houben-Weyl). They are keen that the UK academic community has a chance to evaluate their current online system, which offers the chemist enhanced access to an invaluable source of experimental information. For further details follow the links at

<http://www.thieme-chemistry.com>

We are currently working with Georg Thieme on the best way to coordinate trial access to the system. We hope to contact potential users quite soon, and will provide further details in due course. Contact cdsbb@dl.ac.uk if you have any immediate queries

CDS Road Show and Training

The CDS visited 13 institutions between September '05 and February '06 as part of our Roadshow and training programme. The Service also supported training courses at a number of institutions by providing training Ids and training materials online.

Visit	Site	Date
1	East Anglia Chemical Sciences and Pharmacy	29-Sep-2005
2	Birmingham Chemical Engineering	11-Oct-2005
3	Nottingham Pharmacy	20-Oct-2005
4	Newcastle Chemistry	27-Oct-2005
5	Liverpool John Moores Biosciences	10-Nov-2005
6	Aberdeen Biomedical Sciences	17-Nov-2005
7	Cardiff Chemistry	29-Nov-2005
8	Belfast Chemistry	06-Dec-2005
9	London School of Pharmacy	26-Jan-2006
10	Sunderland Biosci and Chem	08-Feb-2006
11	Cranfield	16-Feb-2006
12	Liverpool John Moores Biosciences	17-Feb-2006
13	Liverpool Chemistry	24-Feb-2006

The CDS website has a large number of training and learning information to enable users to learn more about the Service, data/databases we provide and how to use/better use our databases for their research.

- **Flash Movies**

The website has interactive movies covering Spectroscopy, Organic Chemistry, Physical Chemistry and Crystallography database.

- **Online Documentation Exercises and Tutorials**

- **CDS Overview PowerPoint**

Tailored and more Specific courses

Organic Chemistry:

consists of our hosted reaction and molecular databases. This course also includes demos and set exercises on the use of each one of these databases.

Crystal Structures

Information on the Cambridge Structural Database (CSD), which is the only comprehensive collection of small-molecule organic and organometallic crystal structures. A demo of the software (Conquest, VISTA and Mercury) developed for the search, retrieval, display and analysis of CSD data. A demo of our new web-based crystallographic database tool called CrystalWeb.

Spectroscopy

SpecInfo and Specsurf (a web-based version) - a multi-technique spectroscopic database system, covering NMR, MS and IR spectroscopy

CDS group representatives and departmental heads can choose which one would be best suited for their department. In order to organise a course we require a minimum of six students and you will need to organise access to your local computer training suite facilities. If you are interested please contact: cdsbb@dl.ac.uk

CDS User Forum 2006

The annual CDS user meeting is scheduled to take place on Wednesday the 5th of July at Daresbury laboratory. Topics for discussion will include

- Funding and future of the service
- Roadshow and training
- Database Trials

*The CDS user forum will
take place on 5th of July*

As in previous years the user meeting is an ideal opportunity to meet the CDS team, learn about future plans and to discuss any aspects of the Service. User representatives are particularly encouraged to attend, though all users are welcome. As in previous years, we will be able to cover the travelling expenses of attendees. If you are interested in attending please contact us at cdsbb@dl.ac.uk.

Detherm Update

Detherm 2.1

The Detherm database of thermophysical properties has recently been upgraded to version 2.1. The database has increased in size by about 7%. The number of pure compounds covered has gone up from 24,731 to 25,166 whilst the number of mixtures has increased to 97,046 from 93,505. There are no new client features with this release, though there are some important technical changes. Unfortunately, these changes require users to upgrade to the new version 2.1 client software in order to access the new data. Detailed instructions on how to do this can be found in the download section of the CDS website

<http://cds.dl.ac.uk/detherm21>

Users of the 2.0 client need only install the 2.1 client and make a small change to one of the Oracle client configuration files, whereas users of the older 1.3 and 1.4 clients will need to do a complete install of the Oracle and Detherm 2.1 clients.

Detherm client versions 1.3 and 1.4 are now obsolete, and are no longer supported by Dechema. Any users still using these versions are advised to upgrade immediately, since we intend to switch off the Detherm version 1 server on 18/04/2006.

Users of the version 2.0 client are also encouraged to upgrade in order to access all of the data. We are proposing to terminate the 2.0 service at the end of May this year. If you are likely to have a problem with this deadline, please let us know as soon as possible.

Feedback on the 2005 User Survey

Feedback on the 2005 User Survey

Thanks to all users who took part in the User Survey last year. This generated a lot of useful data which will help us plan future directions for the Service and support the case for continued funding.

The issue of acknowledging and citing the Service in publications generated much comment. We appreciate the argument that the Service can be compared to a library or similar information resource, which researchers would not normally cite in their publications. The reality is that the CDS needs to demonstrate that it is useful for a significant amount of high quality scientific research. Tracking publications is one of the important ways that this can be done. To this end we ask you to think about citing us when publishing work which has benefited from use of the Service.

The Survey produced a number of suggestions for new databases which might be added to the Service. Unfortunately some of these (for either technical or licensing reasons) do not fit in well with our current remote networked service model. In particular systems which are available only as standalone CD-ROM package are in this category. However, we do look into all suggestions, and based on some of them we are proceeding with a number of trials of new systems (see Database Trials article in this Newsletter).

In the context of future development of the Service, the results provided strong support for the retaining our current specialist databases and for the future development of a global molecule database interface. The survey also confirmed that email messages and, to a lesser extent, the Newsletter are the most useful mechanisms for delivering news items.

Finally a number of people commented on the importance of continued support for Macs. We have always attempted to support Macs, though with some of the commercial interfaces this has not always been easy. Recently we have customised the Chemscape system to produce a web version of the Available Chemicals Directory which works without the ChimePro plugin (this is only available for the Windows PCs). This system can be accessed from the CDS home page or directly via the URL

<http://cds.dl.ac.uk/acdmac>

This system is still under development and we welcome user feedback at this stage. We hope to build something similar for our various organic reaction 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.

The above reference is not the *latest* reference describing the CDS. It does, however, allow us to track Service use. A more recent article has also appeared.

"The United Kingdom Chemical Database Service: CDS", Bob McMeeking & Dave Fletcher, in *Cheminformatics Developments: History, Reviews and Current Research* (Ed. J. H. Noordik), IOS Press, Amsterdam, Chapter 2, pp 37-67, **2004**.

Follow the link: <http://cds.dl.ac.uk/cgi-bin/news/disp?citation>

This gives fuller information on acknowledging the Service and its components. It also includes links to the related articles in electronic format.

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 4.7 Million compounds from 24 suppliers.

Crystal 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 CSD.

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,

120,000 proton NMR spectra, 43,000 heteroatom (N-15, O-17, F-19 and P-31) NMR spectra, 140,000 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