Chemical Database Service NEWSLETTER

Autumn 2007

Contents

Welcome

Welcome to the Autumn Newsletter of the Chemical Database Service (CDS)

CDS has operated a successful national chemical database service for the UK academic community for over 25 years. The scope of the provision has changed vastly over the years and the technology used to provide user access to the data has progressed in line with the best appropriate computer hardware, software tools and operational practice.

Inside you will find details of the current state of service and what has changed over the past 6 months.

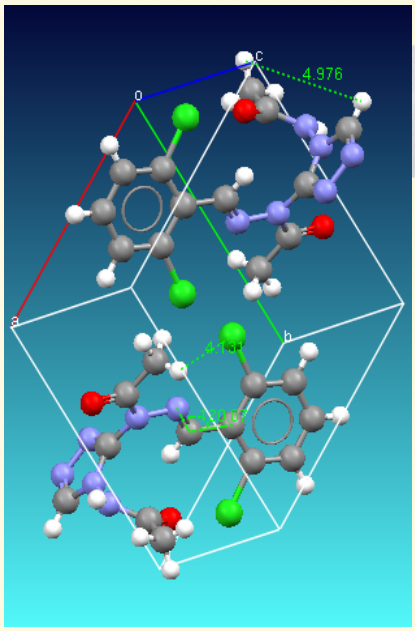
CDS Wins Tender

CDS has won the EPSRC tender to supply chemistry databases to the UK academic community for at least the next 3 years. Details of what the new service will contain after April 2008 can be found on page 2 of this Newsletter and on the CDS web pages

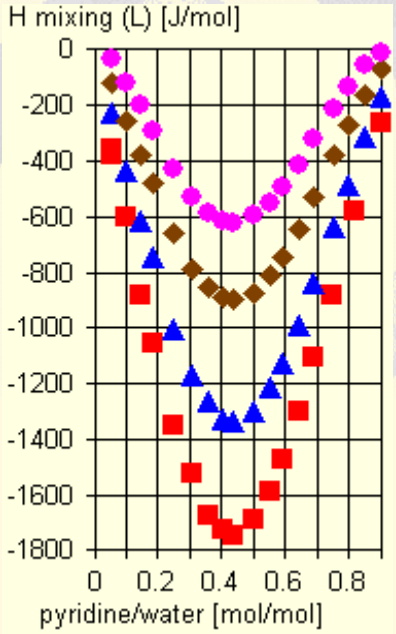
Science & Technology Facilities Council

EPSRC

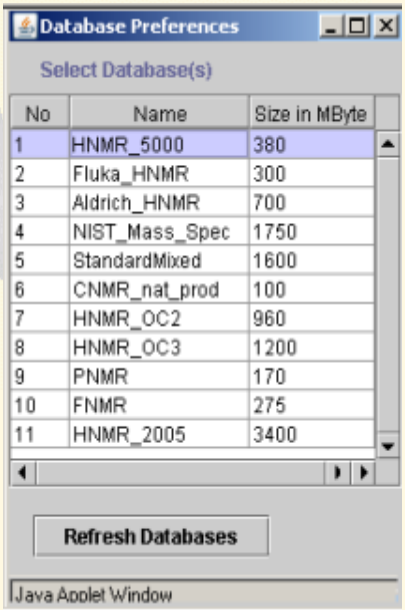
Cambridge Crystallographic Database



DETHERM Thermophysical Database



SpecInfo Spectroscopy Database



Links in this Newsletter can be found at:- http://cds.dl.ac.uk/letterlinks

New for April 2008

Crystallography

ALL presently held crystallography databases will be available and usual updates added.

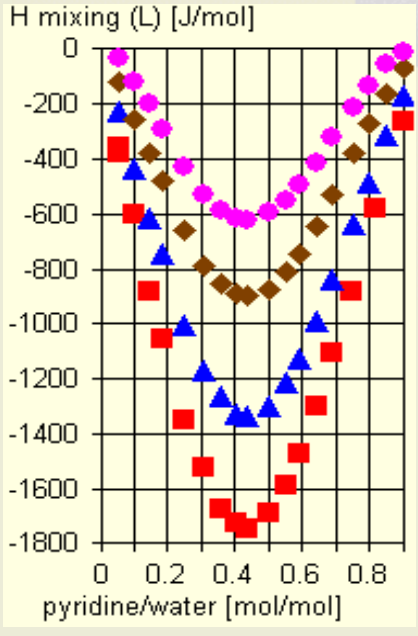
Detherm

New Data Sets

Several new datasets will be added to the collections available via the current service. These new datasets include Gas Solubilities in Electrolyte Systems, Critical Data of Mixture, and Solubilities of large complex compounds.

Web Interface

We will develop a simple web interface which will provide limited searching capabilities on compound names and formulae. It will indicate all the property types that are present in the database for any selected compound but not show the actual data.



Detherm Plot window

ACD/Labs I-Lab System

Following on from a successful trial last year, the ACD/Labs databases will be available for at least 1 year, with the option of extending it for a further 2 years.

Fact File

This comprises a suite of databases including NMR, naming and physical chemical packages.

Data can be obtained by **searching the databases or predicting** using accurate algorithms which can even deal with compounds that exhibit stereochemistry. Formula, name, molecular weight, composition, chemical shifts, coupling constants and references are displayed in a window along with 2D structures plus interactive spectra with zoom facility.

NMR data: Around 375,000 NMR spectra, including:-

- **^{1}H** (over 165,000 structures)
- **^{19}F** (over 13,800 structures)
- **^{13}C** (over 165,000 structures)
- **^{31}P** (over 22,600 structures)
- **^{15}N** (over 8,000 structures)

Naming: The program is able to:-

- Generate a Name according to IUPAC rules
- Generate an Index name according to CAS rules
- Generate a Structure from a chemical name

Physiochemical data: Around 39,000 compounds with including

- pKa (about 16,000 structures)
- LogP (over 18,000 structures)
- Solubility (over 5,000 compounds)

Physical property predictions includes pKa, LogP, LogD, aqueous solubility, boiling point/vapour pressure/enthalpy of vaporization, adsorption coefficient/bioconcentration

SpecInfo

Two new updates (^{13}C NMR and ^{1}H NMR spectra datasets) will be added. These will increase the total number of ^{13}C NMR spectra to 338,000 and the total number of ^{1}H NMR spectra to 183,000.

Authentication

Authentication of users will be handled in two ways. The current CDS id and password system will be retained and will allow access to all components.

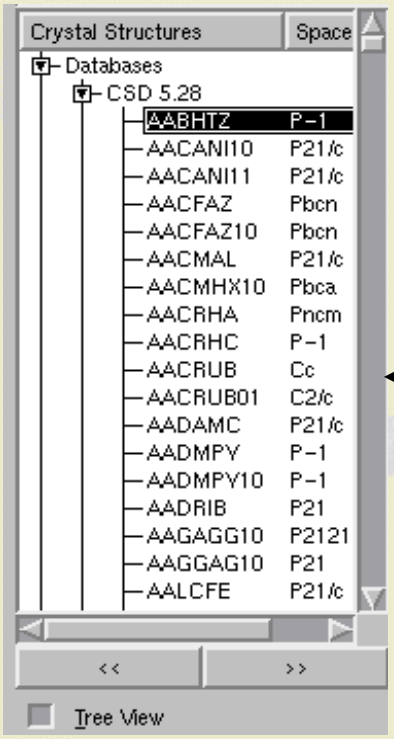
In addition, we will implement Shibboleth based authentication for the web interfaces, which will allow members of the UK Access Management Federation for Education and Research to authenticate with their Institutional credentials (usually id/password). This system should also be compatible with Athens id/passwords, for Institutes which are still using it.

New Features

Crystallography

The **CCDC** 2007 release contains:

- The Cambridge Structural Database, Version 5.28
- ConQuest 1.9
- Mercury 1.5
- IsoStar 2.0 (and IsoStar 2.0 PC Client)
- Mogul 1.1.1
- Vista
- PreQuest



Mercury Tree View

New Features

ConQuest

New options are available when exporting CIF files.

Mercury 1.5

There have been a number of improvements such as:

- The scrolling list of structures on the right-hand side of the Mercury window has been replaced by a tree-view
- Various improvements to simulated powder diffraction patterns

(for a full list of improvements, see the link)

IsoStar 2.0

Scatterplot data has been updated and now includes 21,850 scatterplots from CSD 5.28, 7,129 scatterplots from the PDB and 1,550 potential energy minima.

ICSD-WWW

- The ICSD-WWW interface has been upgraded. A major difference is the availability of a new Jmol applet option which allows platform independent display of structures.

Jmol

- CrystalWeb structure display facilities are now available from within the CDS implementation of ICSD-WWW. Thus it is possible to generate hits using ICSD-WWW search facilities and then create CrystalWeb displays. To allow this a "Cweb" buttons has been added to the main ICSD-WWW hits window and the "Details" sub-window.

Cweb

CrystalWeb

A new coordination search form is now available. It searches the ICSD and CrystMet databases and allows searching on a specific atom type and coordination number.

Links from this page and others can be found at:- http://cds.dl.ac.uk/letterlinks

Current State of the Service

```
Welcome to the UK Chemical Database Service on
cds.dl.ac.uk
Unauthorised access is not permitted.

cxxx@cds.dl.ac.uk's password:
Last login: Thu Oct  4 10:15:01 2007 from 111.123.45.100
*************************************************
* 18/09/07  The August 2007 update for the Cambridge    *
*           Structural Database has just been received and *
*           is now available via ConQuest and  Quest. This  *
*           update adds 8,500 structures.               *
*************************************************
 NOTE: You are now logged onto the new server. You should
notice no significant difference as compared to the old system
other than a faster response. However, if you suspect any
problems contact cdsbb@dl.ac.uk.

For the complete Help and News details check the CDS Web
                    (http://cds.dl.ac.uk)
cds%
```

Databases

All databases in the

- **Crystallography**
- **Spectroscopy**
- **Thermophysical**

areas are available as before and will continue until April 2008. The only difference is the fact that the organic chemistry databases (reactions and available chemicals) have had to be removed due to EPSRC withdrawing funding.

New Machine

On Monday 17th September 2007 there was a major upgrade to the Service. We replaced the current stopgap machine which has been hosting the Service since the failure of the original server over the summer. The new server is a substantially more powerful machine than either of the above. Whilst the upgrade should have been mostly transparent to users, there are a few points to note: The new machine uses the same alias (cds.dl.ac.uk) as the old, though the actual address is different. The new server is named cds6.dl.ac.uk with IP address 193.62.124.56. However, it is best to use the alias **cds.dl.ac.uk** for access. If you used the old address (cds8.dl.ac.uk) for access, you will need to change this to cds.dl.ac.uk.

Problems Accessing the New Server

A few people have had problems with accessing the new CDS server. Virtually all of these are due to the fact that the new server has a different name from the old one.

X-Windows Access

X-Windows access may be a problem, due to xhost permissions or site firewalls. Typically, when you try to open a window (such as when you start ConQuest) you will get an error of the form:

Error: Can't open display: .....

a) PC Users

If you are using eXceed on a PC you may have 'Host Access Control List' set to File which usually points to a text file called xhost.txt. You should edit this file to include cds.dl.ac.uk.

b) Unix/Linux users

If you are using a Unix/Linux box, make sure you type xhost cds.dl.ac.uk before logging on to CDS.

c) Firewalls

Many site firewalls only allow X-Windows to be sent from specific IP addresses. If this is the case you will need to get you IT support to add the new machine to this list.

SSH

Most of these problems can be avoided by using SSH (Secure Shell) and port forwarding. Details of how to set this up for a PC are given here and for a Mac/Linux here (or go to the Newsletter links web page)

DNS spoofing

If you use ssh, you may find that you now get a warning or error concerning "DNS spoofing" when you try to connect (mainly Mac and Linux users).
This is because the RSA host key for cds.dl.ac.uk has changed, as it is now a new machine. This problem can be fixed by going to the .ssh directory under your home directory and editing the known_hosts file with your text editor. You will need to remove the line which starts cds.dl.ac.uk.

Links from this page can be found at:- http://cds.dl.ac.uk/letterlinks

Rationalisation of Services

Access Method	Packages
WEB	ICSD CrystalWeb Isostar Specsurf
X-WINDOWS	Conquest Isostar Mercury VISTA Mogul
Client/Server	Detherm Isostar
Telnet	Removed
Secure Shell (SSH)	Use with x-windows packages

```
Welcome to Microsoft Telnet Client

Escape Character is 'CTRL+1'

Microsoft Telnet >open cds.dl.ac.uk
Connecting To cds.dl.ac.uk ....
Could not open connection to the
host, on port 23:
Connect failed
```

Removal of Telnet Access

The telnet access service has been withdrawn. The few users who still used this method of access to CDS will need to switch to using secure shell (SSH). Details are available on the CDS web site, or contact us at cdsbb@dl.ac.uk if you have difficulties.

Removal of Users

```
login as: cxxx
Welcome to the UK Chemical
Database Service on cds.dl.ac.uk
Unauthorised access is not
permitted.

cxxx@cds6.dl.ac.uk's password:
Access denied
```

User ids and their files were automatically transferred to the new machine. However, the following categories of user were not transferred:-

- Users reported as having left by their departmental rep.
- Those users whose email address is no longer valid.

Our intention is only to remove ex-users who have moved on.

If your CDS Id/password are not accepted, it is possible that your account was not transferred to the new server (accounts that we believed to be no longer valid). Please contact us and we will sort it out.

Old Command Line Codes Removed

```
cds% cssr
cssr: Command not found.
cds% icsd
icsd: Command not found.
cds% mdf
mdf: Command not found.
cds% cdif
cdif: Command not found.
cds%
```

The old crystallographic command line codes CSSR, ICSD, MDF and CDIF were not transferred to the new machine. These codes only work with old copies of the data (at least 2 years out of date) and we cannot now update them. They have largely been superseded by more modern packages. However, if their loss will cause major disruption please contact us. All other databases and programs should work exactly as before. However, if you encounter any problems with accessing or suspect problems with any of the packages please contact us at cdsbb@dl.ac.uk.

Electronic Literature Linking

CrystEngComm, 6 (2004) p633

doi: 10.1039/b415627g

LitLink is no longer available for literature linking in ConQuest or CrystalWeb. This function is now handled using the **CrossRef** system, or directly with Digital Object Identifiers (**DOI**s) where they are available.

Links from this page can be found at:- http://cds.dl.ac.uk/letterlinks

Removal of Services

Since April 2007, the CDS has received no funding to provide data in the Organic Chemistry area. This has resulted, not only in the loss of these important sources of information, but also some loss of funding for personnel to run the service.

Since the announcement last year, we have tried to re-establish these key databases. We were fully aware that their loss has a huge negative effect on UK research work. We, therefore, arranged stop gap provision for a further two months with the hope and expectation that suitable arrangements could be arranged for the community in the longer term.

We have held meetings with EPSRC, JISC and CHEST representatives to see if alternative arrangements might be possible. JISC were willing to pursue this route, recovering costs via a site subscription mechanism in many ways similar to that already in place for Beilstein/CrossFire, etc.

The EPSRC were also very receptive to the idea, but were unable to come up with any additional funding which could have carried us over the transitional period. In this context negotiations with our current main database and software supplier foundered and agreements for a replacement service have so far not been possible.

The following databases have been lost under the new arrangement:

REFLIB (Reference Library) - consisting of :-
- **THEILHEIMER** (Synthetic Methods)
- **CLF** (Current Literature File)
- **CHIRAS** (Asymmetric Synthesis)
- **METALYSIS** (Metal-mediated transformation)
- **ORAC CORE** (Established Literature)
- **CHC** (Comprehensive Heterocyclic Chemistry)

REACCS-JSM (Journal of Synthetic Methods)
CHEMINFORM (CIRX - current awareness)
ORGSYN (Organic Synthesis)
SPG (Protecting Groups)
SPS (Solid Phase Synthesis)
BIOCATALYSIS (Enzymes etc. as catalysts)
CHIRBASE (Chiral Separations by chromatography)
NCI (National Cancer Institute Database)

ACD (Available Chemicals Directory)
SCD (Screening Compounds Database)

Alternatives?

MDL Databases

Most of the above databases can be accessed via MDL's DiscoveryGate. This would require a site licence fee from each University similar to the Beilstein CrossFire deal.
Catalogue information can often be obtained from the web site of individual suppliers. See the CDS external links web pages for a list of Chemical Sources

Accelrys Databases

SPG, SPS and BioCatalysis can be accessed using Accelrys' Accord software (Chemistry Cartridge and Database Explorer software plus Oracle database required)

ChirBase

ISIS/Base desktop or ChemAxon JChem versions may be available, or ChirBase is available for ACD/Labs ChronManager.

The CDS is still investigating alternative ways of providing access to these or equivalent databases.

Links from this page and others can be found at:- http://cds.dl.ac.uk/letterlinks

CDS Research Highlight

The Oxidative Spirocyclisation of 2-(ω-Hydroxyalkylfurans) as a Key Step in Natural Product Synthesis

Jeremy Robertson **(jeremy.robertson@chem.ox.ac.uk)**

Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA

During the development of our synthetic approach to the lituarines1,2,3 (**Box 1**) we recalled that the oxidation of substituted furans with epoxidizing agents results in ene-diones. Application of this reaction to furans bearing a hydroxyalkyl group at the 2-position was expected to lead to a double cyclisation process resulting in spirocyclic butenolide products. In reducing this idea to practice, we were pleased to find that this proposal worked well using an excess of MCPBA at room temperature. In the context of our lituarine synthesis, this process led to the rapid assembly of a large fragment of the natural products, the C(6)–C(18) tricyclic spiroacetal (**Box 2**).

Soon afterwards, we discovered that these spirobutenolide intermediates were viable substrates for conjugate addition with *S*-stabilized organolithium reagents and, at low temperature, addition proceeded kinetically *anti*-to the oxygen in the tetrahydropyranyl ring; this sense of stereoselectivity matched that required for the methyl group in the lituarines.

Further elaboration of the conjugate adducts allowed introduction of the C(18)–C(19) bond bysilyl enol ether addition to an intermediate spiroacetal oxonium ion, an unprecedented transformation with wide scope for application in natural product synthesis (**Box 3**).

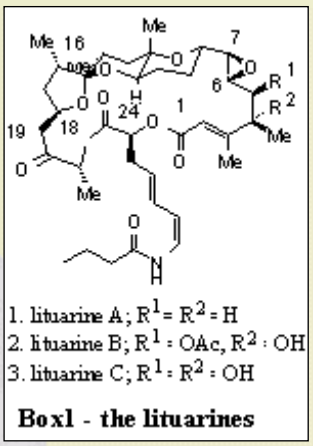
Collectively, these results constitute a powerful general approach to a variety of natural products based on substituted spiroacetal structures and we have completed a number of these targets and are actively pursuing others.

As specialists in synthetic methodology and natural product synthesis, we are engaged in searching on a daily basis for literature precedent for specific transformations and the reaction datasets hosted by the Chemical Database Service are indispensable. Although these datasets are not comprehensive, the flexibility available in setting up reaction searches is unparalleled which results in the highest quality hits. We greatly value this service and we very much hope that it will remain accessible to the academic community.

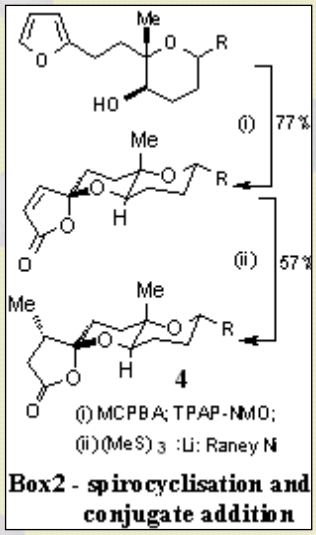
(1) J. Robertson, J. W. P. Dallimore, P. Meo, *Org. Lett.* **2004**, *6*, 3857–3859.

(2) J. Robertson, P. Meo, J. W. P. Dallimore, B. M. Doyle, C. Hoarau, *Org. Lett.* **2004**, *6*, 3861–3863.

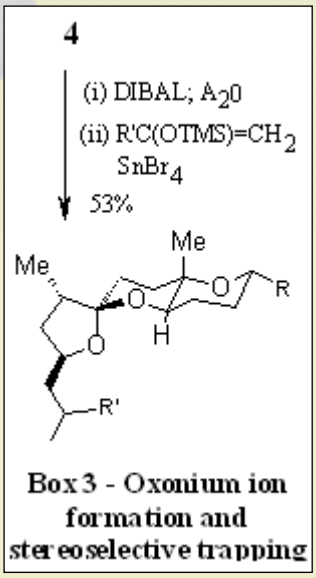
(3) J. Robertson, J. W. P. Dallimore, *Org. Lett.* **2005**, *7*, 5007–5010.



Box1 - the lituarines



Box2 - spirocyclisation and conjugate addition

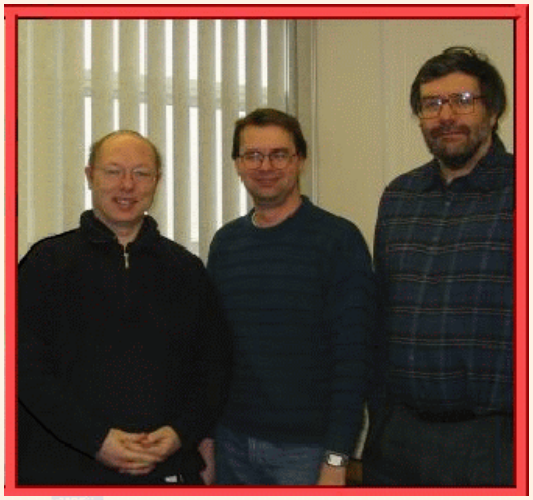


Box 3 - Oxonium ion formation and stereoselective trapping

The CDS Team

Who Are We?

The CDS group currently operates as a well integrated team of three full time staff who are all post-doctoral chemists with around 50 years of experience in chemistry database provision between them. Currently, and over the next 3 years, only two members are fully funded.



Dr. Don Parkin **Dr. Dave Fletcher** **Dr. Bob McMeeking**

Help Desk

A help desk, accessible by phone or email, is available to answer all queries and there is extensive on-line help through web pages, which include user guides, FAQ database, tutorials and exercises as well as Flash Movies that show what the different packages can do. There is also a JISCmail discussion list available to users. Site visits and training courses are available on request, at no cost to the user (see training section below).

CDS has site representatives at most Universities and a number have their own training IDs to help run courses internally.

Training

The CDS web site contains online training material for all of the databases and interfaces that are available. This includes “getting started” and tutorial information as well as structured problem sets, appropriate Flash Movie clips, etc. This will be augmented and updated in line with system developments.

The service offers “hands-on” training courses at University sites on demand. These require the site to have a suitably equipped training room as well as sufficient demand (minimum of 6 people) for a course. Individual sites will not be charged for these services.

User Meeting

We continue to run an annual user meeting and produce a newsletter twice a year.

Where Are We?

We are part of the support and service programme of the Computational Science and Engineering Department (CSED) based at Daresbury Laboratory, which is in turn now part of the Science and Technology Facilities Council (STFC). The CDS is one of the National Services funded by the Chemistry Programme of the Engineering and Physical Sciences Research Council (EPSRC).

Daresbury Laboratory

Links from this page and others can be found at:- http://cds.dl.ac.uk/letterlinks

Databases Available From The Chemical Database Service

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

STRUCTURES

The Structures databases contain a comprehensive collection of organic, organometallic and inorganic compounds, metals,and alloys **crystal structure data**.

CSD - Cambridge Structural Database.

Crystal structure data for over **423,000 organic and organo-metallic compounds**. New releases of this database are received and mounted bimonthly.

* **ConQuest** is the graphical front end which has 3D search capabilities.
* **VISTA** performs numerical, statistical and graphical analyses.
* **Mercury** provides comprehensive facilities for visualising crystal structures in three dimensions.
* **IsoStar** A knowledge base of non-bonded interactions derived from the CSD, the Protein Data Bank (PDB) and molecular orbital calculations. Uses a simple web interface.
* **Mogul** A knowledge base of molecular geometries using data derived from the CSD.

ICSD - Inorganic Crystal Structure Data File.

Crystal structure data for Over **97,000 inorganic compounds**

WWW interface available.

CRYSTMET

Crystal structure data for over **115,000 metals, alloys and intermetallics.**.

CDIF - Crystal Data Identification File.

Crystal class and unit cell data for **237,671 crystal Structures**.

CrystalWeb

A simple **web interface to all of the crystallographic databases** that allows bibliographic and cell data searching along with structure display.

SPECTROSCOPY

The Spectroscopy databases are designed to aid the chemist in structure elucidation and spectra interpretation problems.

SPECINFO

SpecInfo is a multi-technique spectroscopic database system which covers NMR, IR and mass spectra. A variety of features are available within the program to help with spectrum prediction and searching. It contains spectral data sets with their associated structure connection tables. The database currently contains:-

^{13}C NMR	^{15}N NMR	^{17}O NMR	^{31}P NMR	^{19}F NMR	^{1}H NMR	IR Spectra	MASS Spectra
102,369	992	856	16,561	25,442	117,379	20,898	138,727

SpecInfo is accessed using **SpecSurf**, a **Web graphical user interface** that makes drawing structures, creating peaklists and viewing hit lists easy.

PHYSICAL CHEMISTRY

DETHERM

One of the world's largest thermophysical property databases of pure compounds and compound mixtures. Contains nearly **6 Million data sets** for around **127,000 systems** (around 26,500 pure substances and 101,300 mixtures) covering around **500 property fields**.

UTILITY PROGRAMS

A variety of utility programs are available, including links to electronic literature, chemical file format conversion and molecule viewers.

CRAD - A crystal radial distribution calculation program.

BABEL and **BEDLAM** - file format converters

General Information

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

The Service is available **free of charge at point of access** 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 ID's.

Main Service Machine

cds.dl.ac.uk

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

World Wide Web Site

Information about CDS, including online help, documentation and up-to-date news items is available over the World Wide Web at the CDS website.

http://cds.dl.ac.uk

Web based interfaces to selected Databases are also available from this site.

Documentation:

Most documentation is available online or within the client packages. As well as User Guides, training exercises and tutorials can also be downloaded from the CDS web site.

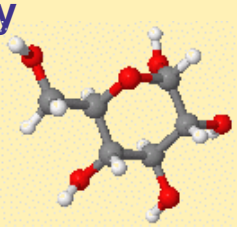
Comments:

All comments, questions and suggestions about this newsletter should be sent to:

Dr. D. Parkin
Email: D.Parkin@dl.ac.uk
Phone: 01925 603 162

Links within this Newsletter can be found at:
http://cds.dl.ac.uk/letterlinks

CHEMICAL DATABASE SERVICE
Daresbury Laboratory
Warrington
Cheshire
WA4 4AD
Tel: 01925 603 162
Fax: 01925 603 031
Email: cdsbb@dl.ac.uk



CDS
ON-LINE