Chemical Database Service

NEWSLETTER

Issue 11 October 1999

A three month trial is underway of a number of life science related databases accessible through the ISIS software. Once the trial is complete, users will be asked whether any of the databases are suitable to be included in the permanent portfolio available from CDS. Details of the databases available in the trial are given in the article on page 2.

A variety of new data has become available in the past few months. This includes new proton NMR and mass spectra for SpecInfo (see page 5) as well as updates to the ChemInform reaction database (see page 7).

All ISIS users will need to update their local Reaction Browser file in order to access the new ChemInform data. Details of how to do this are given on page 7.

SpecSurf is the web-based interface to the SpecInfo database. Currently we have a beta version of this interface available for users to try out. See page 8 for more details. We would welcome feedback on any aspect of the interface, but particularly about ease of use and complexity.

We should have a test application for Chemscape, the web-based interface to the ISIS databases, available next month for users to try out.

CDS is actively considering the Detherm database in order to expand the physical database portfolio. The article on page 5 gives more information on the scope and content of the database. Could anyone who might be interested in helping us evaluate this resource please contact us.

Users are reminded that CDS offers training courses in all of the databases, free of charge, at your site. In addition there is a wealth of training material available on the CDS website at http://cds.dl.ac.uk/cds.

Contents

The Chemical Database Service is funded by the Engineering and Physical Sciences Research Council

Trial ISIS databases

The value of chemistry as the classical science underpinning progress in the life sciences is increasingly getting its deserved recognition. Accordingly we have set up an evaluation trial for a number of databases related to biomolecules and their analogues. The trial will run for three months, starting in September 1999, at the end of which a questionnaire will be sent out to obtain user's views on the usefulness of the databases and whether they should be added to the list of permanently available databases.

To evaluate the databases, download the remote access browser files from the CDS Web site at **http://cds.dl.ac.uk/cgi-bin/down/isis.html.**

The following databases have been selected. More details on each can be obtained from the CDS website at http://cds.dl.ac.uk/cds/ISIS/trial/trial.html.

BioCatalysis (from Synopsys)

This is a database of biocatalysts which can be used to effect novel transformations and as versatile replacements for traditional catalysts in organic synthesis. The database offers a number of advantages:

High chemo-, regio- and enantioselectivity.
Novel transformations become possible.
Ecologically sound - non-toxic biomass, energy efficient.
Effective under mild conditions - ambient temperature & pressure, neutral pH, aqueous and inert solvents.

Database Summary

Source	Thorough coverage of the scientific literature, including patents, from 1903 to the present.
Production	Developed in collaboration with academic experts in the field, Professors Bryan Jones (Toronto) and Herbert Holland (Brock).
Focus	Synthetic use of enzymes and micro-organisms.
Size	Version: 99.1 contains 26,479 reactions; 18,583 citations
Maintenance	Six-monthly, totaling approximately 2,000 reactions per year

Browser File and Forms

The database can be remotely accessed via the Reaction Browser (version2). This means that :

It can be added to the list of available reaction databases

Searches can be conducted over specific datafields with a list of all the available data for that field.

Results can be clustered.

Metabolism (from Synopsys) and **Metabolite** (from MDL)

An understanding of the likely metabolic fate of a compound can help predict its biological activity, toxicity and its effects on the environment.

Both databases cover the details of the metabolic fate of compounds. They comprise biotransformations of a range of compounds including pharmaceuticals, agrochemicals, food additives, and environmental and industrial chemicals.

Metabolism

Database Summary

Source	Biotransformations, Volumes 1-7 (RSC: Ed. David Hawkins)
Production	Database compilation and quality control is carried out by Synopsys.
Focus	Metabolism of drugs, agrochemicals, food additives, environmental and industrial chemicals in vertebrates.
Size	Version 1.0 (August 1997) comprises 9,300 transformations of 1,916 parent compounds, abstracted from over 1,900 primary literature sources.
Maintenance	The database will be updated periodically as new material becomes available. Augmenting the database with additional material is also under consideration.

Browser File and Forms

The browser files consist of 2 main forms - MB1 and MB2 - a default 'reaction' form and a molecule form. The database is searchable by exact and similar structure, substructure, text and keywords.

Metabolite

Database Summary

Sources	Biotransformations of Drugs (1977-1983); Pharmacokinetics (1986-1990); Original metabolism literature and new drug applications (1990-present); Proceedings from International Society for the Study of Xenobiotics (ISSX) meetings.
Focus	Xenobiotic compounds and biotransformations (primarily medicinal drugs), indexing of path and scheme information; experimental data from in vivo and in vitro studies.

Size	Version 99.1 contains 115,900 transformation records; 7,600 parent compounds; 46,000 transformations; 30,500 molecules.
Maintenance	1901-present (15,900 references); updated semi-annually.

Browser File and Forms

Searches can be conducted using three different forms - Molecule, Transformation or Data. There are also two sets of five different forms for viewing the hits.

Two buttons are available down the left-hand side to switch between the two sets of forms.

Transformation and Schemes button

Molecules button.

Bioster (from Synopsys)

Bioster is a database of bio-analogous molecule pairs (including drugs, agrochemicals and enzyme inhibitors). Replacement of bioisosterically equivalent moieties is an important and well recognised approach to help optimise activity, remove unwanted size effects, or change a bulk property such as the transportability of a drug or agrochemical.

Database Summary

Sources	Primary journal articles, review papers and books during the past 35 years up to early 1998.
Production	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.
Focus	Bio-analogous molecule pairs (including drugs, agrochemicals and enzyme inhibitors).
Size	Version 99.1 contains 4,932 molecule pairs; Over 7,500 active compounds; 8254 literature citations.
Maintenance	1907-1998; updated annually.

Browser File and Forms

Three forms are available for searching and viewing the data. The main reaction form (BS1), the molecule form (BS2) and the fragment form (BS3).

New thermophysical data

We intend to expand our physical data portfolio - an area which is not currently well represented in CDS. In particular the Detherm database is under consideration.

Detherm, from Deutsche Gesellschaft für Chemisches Apparatewesen, Chemische Technik und Biotechnologie e.V. (DECHEMA), contains thermophysical data, much of which has great value in the design of various processes in the chemical industry.

The database contains some 3.35 million data sets for around 105,000 systems (18,000 pure substances and 87,000 mixtures) covering more than 500 properties, encompassing thermodynamics, phase equilibria and electrolyte data. Details are available on the CDS website or the Detherm website, which is linked from there.

We currently have an evaluation PC version of the "Inhouse" interface which has full functionality but with a representative sub-set of the full data. We can arrange to make this available for external testing. If you are interested please contact r.f.mcmeeking@dl.ac.uk.

If you wish to get a better feel for the scope and quantity of the full data, you can perform full searches over the WWW from the Dechema site and get details of the size of your hits sets. This is free of charge but you would need to pay to download the actual data.

We anticipate these data being of interest to a significant number of people who are not traditional users of CDS (many physical chemists as well as chemical engineers). If you have any colleagues who may be interested, please inform them.

We are also looking for other suitable physical data sets. If you have any suggestions, please let us know.

New SpecInfo data

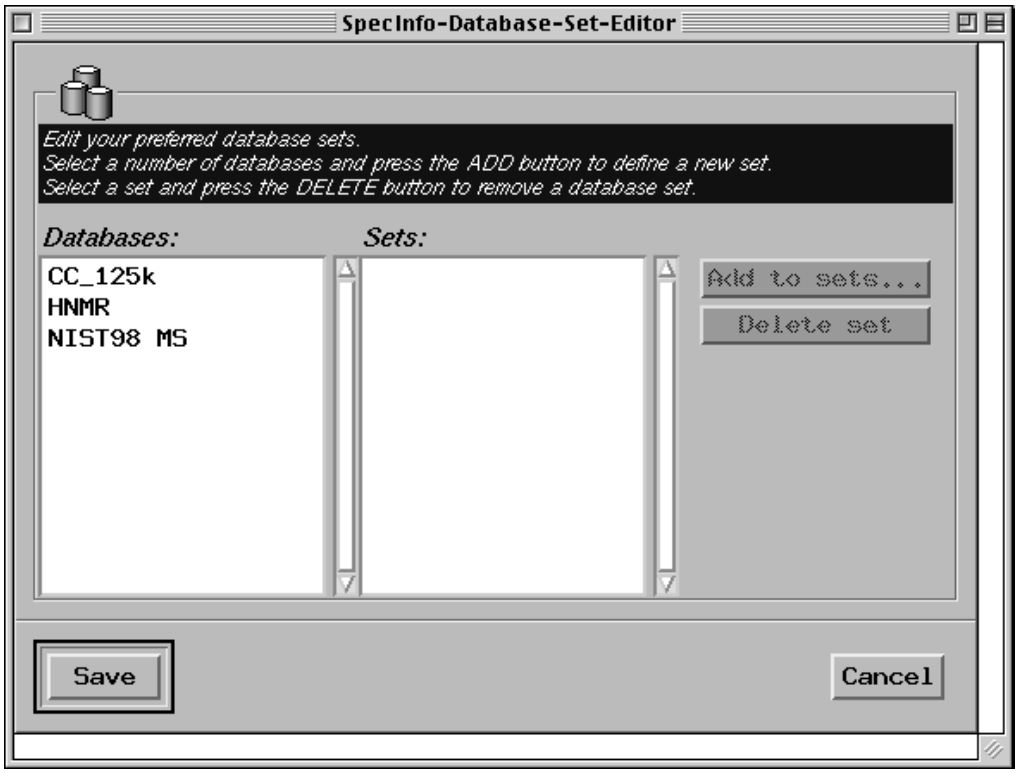
Two new datasets have been added to SpecInfo recently:

1) An additional 2,500 spectra proton NMR dataset is now available, bringing the total number of proton NMR spectra available to 5,000, of which 3,200 are fully assigned. These spectra can be compared with experimental data and the associated bibliographic and structural data are fully searchable. Selected spectra can also be output in the JCAMP-DX format.

2) The NIST / HIH / EPA mass spectra dataset has been added, containing 128,000 mass spectra and associated structures. These spectra can similarly be compared to experimental data and output in the JCAMP-DX format.

A new tutorial guide to using SpecInfo (called the Illustrated Guide) can now be viewed on the CDS website. This covers most features of the program, including structure drawing, spectra and structure searching and spectrum prediction.

Now that there are several databases available within SpecInfo, you may need to run a search in more than one database at a time. To do this you must define a set of databases and run your search in that. To define a set of databases, select the Edit Database Set option from the Edit menu of the Mainwindow. This brings the dialog box shown below.



Hold down the Shift key and click on the databases that you want in the set. Click on the Add to sets button and you will be prompted to enter a name for this set of databases. Once you have saved this set you will need to disconnect (Disconnect option from Connect menu) and relogin (Login option from Connect menu) to SpecInfo to use it.

Now, when you are in the query editor, you can select any of the databases or sets of databases for searching from the Databases menu.

ISIS - are you searching all of the data?

The ChemInform RX database currently contains over 575,000 reactions and is growing at the rate of 70,000 reactions per year. This is the main reaction database searched via ISIS. In order to access the new data added each year, users need to modify their local ISIS database access file or download a new version from us (see below for details). If you do not do this, you may not be searching the most recent reaction data.

Statistics for August 1999 show that 2% of all searches in the ChemInform database were not accessing the data added in 1997 and nearly 30% of searches were missing the data added last year.

This year's update to the ChemInform RX database has just been installed (September 99), and ALL users will need to update or redownload their database access files to use this data.

How to check whether you are accessing all of the ISIS data

- Start ISIS/Base.
- From the File menu select Open Database.
  Mac Select the RXN Browser file from the RXN Browser folder.
  PC Select the file rxnbrows.db from the rxnbro directory.
  SGI Select the file all.db.
- Click on the expert query box (Mac and PC only).
- Select Retrieve All from the Search menu.
- ISIS should retrieve a hitlist of 896,722 reactions. If you get fewer reactions you are not accessing all of the data available.

How to update your ISIS database access file

If you run ISIS/Base from a networked server, you may need to get your network administrator to perform the following steps.

- Start ISIS/Base.
- From the File menu select Configure Database (if you already have the database open, you will need to close it before you can configure it).
  Mac Select the RXN Browser file from the RXN Browser folder.
  PC Select the file rxnbrows.db from the rxnbro directory.
  SGI Select the file all.db.
- Check the large box, it should contain the following list of 15 hviews (the order is not important) :

cirx2000+cirx99+cirx98+cirx97+cirx96+cirx95+cirx94+cirx93+cirx92
+reflib+jsm+chc+spg+sps+orgsyn

- If any of the database names are missing, edit the box, then click on OK.
- Your database access file has now been updated.

How to download a new ISIS database access file

You can download new ISIS database access files from the CDS website at http://cds.dl.ac.uk/cds. Choose the Download link and following the procedure for downloading ISIS, but once you are on the ISIS download page you only need to get the new database access file. The appropriate link depends on your computer type as shown below.

PC:
CDS Browser Files link - to update if you already have the Browser.
Reaction Browser 2 link - for full Browser install.
Mac:
RXN Browser2 link.
SGI:
Main reaction databases link.

We maintain up to date online information of the status of reaction databases accessible via the Reaction Browser (ISIS). Type "browser_news" when logged on to the CDS server to view this information, or look at the news link on the CDS website.

SpecSurf - a web interface for SpecInfo

A beta version of SpecSurf (the new web interface for SpecInfo) is now accessible from the CDS website. This interface allows users to perform a subset of the SpecInfo functions using a Java enabled browser. The functions available include:

Spectrum prediction
Structure and sub-structure searching
Bibliographical information searching
Limited spectrum searching

Features not currently available include most of the spectrum searching modes as well as the ability to backtrack to the original data used in a spectrum prediction.

The interface can be accessed from the CDS website, or directly at the URL http://cds.dl.ac.uk/specsurf. To use SpecSurf you need your CDS id and the password specinfo (in lower case).

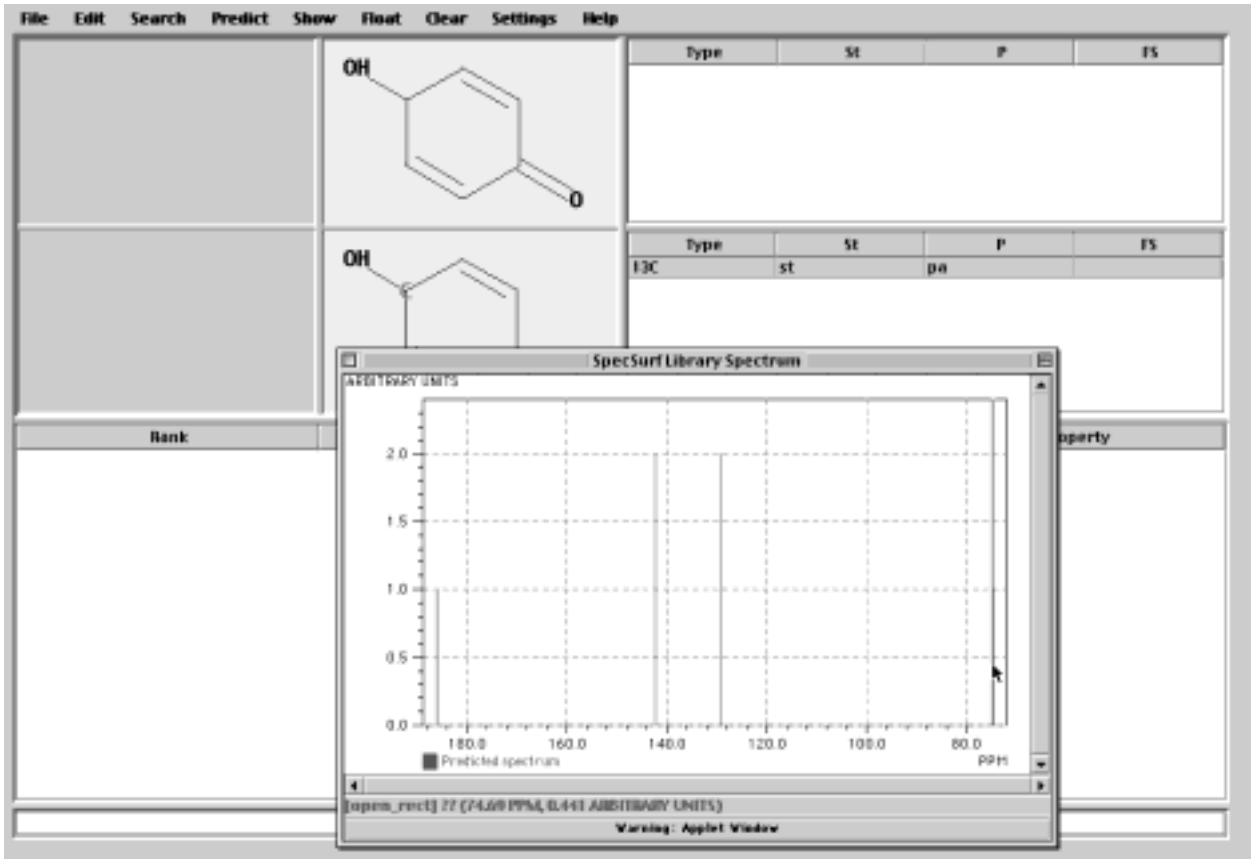
The interface is currently a beta release and therefore may well contain significant bugs. If you have any comments about the interface, including bug reports, please send them to d.a.fletcher@dl.ac.uk.

We have found that the interface currently only works for a limited set of browsers. Browser/platform combinations which we have found to work include:

PC (NT 4) with Netscape Communicator 4.61
PC (NT 4 or Win95) with Netscape Communicator 4.05 and the Javasoft Java Plugin.
Macintosh (OS 8.1) with Internet Explorer 4.5 using the Apple MRJ 2.1.4 Java Virtual Machine.

For more details of supported browsers, see the initial SpecSurf page at the URL given above.

A view of the SpecSurf interface showing a carbon NMR spectrum prediction is shown below.



Secure Shell

Secure Shell provides encrypted telnet-style sessions between your terminal and CDS. In particular, your id and password are encrypted before being sent across the network so you are not vulnerable to "sniffer" programs. Go to the following URL for more details, http://www.cs.hut.fi/ssh.

Secure Shell consists of two parts, a server (the sshd daemon) which runs on cds, and a client which runs on your own machine. Free clients are available for most platforms.

PC - TTSSH
http://www.zip.com.au/~roca/ttssh.html
Mac - NiftyTelnet
http://andrew2.andrew.cmu.edu/dist/niftytelnet.html
UNIX - ssh
http://www.cs.hut.fi/ssh

The UNIX client can also be setup to allow X-Windows to pass through the encrypted channel. This can be useful to allow X-Windows access from sites where xhost based authentication has been blocked.

Protein Data Bank (PDB) update

The PDB contains structural information about proteins. The database is growing at a significant rate, though it is currently little used on CDS. This is probably because the search tools available on CDS have been superseded by features that can be accessed on the PDB website or one of its mirrors. In addition Cambridge have now dropped support for PDB searches via the QUEST software. However the use of derived PDB nonbonded interactions in IsoStar continues to be fully supported.

In view of these facts, we intend to drop support for the PDB, though we will still retain appropriate links from the CDS website. To access the PDB in the future, you should go to one of the UK mirrors of the PDB site. These are listed below.

http://www2.ebi.ac.uk/pdb
http://pdb.ccdc.cam.ac.uk/

Summary of databases available from the Chemical Database Service

ISIS A chemical reaction information management system allowing search, retrieval and display of molecules, reactions and their associated data. Currently contains 897,000 searchable reactions. Of particular interest are the various specialist databases which we support. They currently include Protecting Groups, Solid Phase Synthesis and Chiral Separation data.

In addition there is the Available Chemicals Directory (ACD) which contains supplier information on 270,000 different compounds from over 500 different suppliers. This is likely to be of value to anyone who has a need to order chemicals of any description.

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 ^{13}C NMR; 999 ^{15}N NMR; 856 ^{17}O NMR; 2,183 ^{31}P NMR; 1,825 ^{19}F NMR; 5,000 ^{1}H NMR, 20,898 infra-red spectra and 128,000 mass spectra.

CSD The Cambridge Structural Database. Crystal structure data for 198,000 organic and organometallic compounds. Currently available under both Cambridge's QUEST retrieval software and our in-house software (CSSR). The QUEST software has a number of features not available in CSSR, such as 3D searching.

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

ICSD Inorganic Crystal Structure Data File. About 50,500 inorganic structures - the companion file to the Cambridge organic file. ICSD is now available via a WWW interface.

MDF Metals Data File. Crystal structure data for nearly 56,000 metals, alloys and intermetallics.

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

FNMR A databank of 6,000 ^{19}F NMR spectra and coupling constants.

ELYS Electrolyte Solutions Database. Thermodynamic and transport property data such as density, viscosity and diffusion coefficients. Currently contains about 10,000 entries.

Information

General queries:

Computer Help Desk / User Interface Group (UIG)
Daresbury Laboratory
Daresbury, Warrington
WA4 4AD
(01925) 603351
Email uig@dl.ac.uk

Hours are:
Mon-Thur 9:00-12:00, 13:00-17:00
Fri 9:00-12:00, 13:00-16:00

Otherwise 24 hour answer phone.

General database queries:

Chemical Database Service (CDS)
Room C18, Daresbury Laboratory
Daresbury, Warrington
WA4 4AD
(01925) 603162
Email cdsbb@dl.ac.uk

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) 603162

For QUEST specific problems:

Dr. R. F. McMeeking
Email R.F.McMeeking@dl.ac.uk
(01925) 603162

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

Registration of new accounts:

You can now register online for a CDS account from the CDS website.

Documentation:

Most documentation is available online and some can be downloaded from the CDS web site. Printed manuals can still be ordered from UIG at the above address.

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.

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:

Dr. D.A. Fletcher
Email D.A.Fletcher@dl.ac.uk