



---

# Chemical Database Service

## NEWSLETTER

Issue 9

June 1998

---

The new CDS server is now online and available to users. See page 2 for more details and for instructions for moving your account to the new machine.

A new version of the **ISIS Reaction Browser** is now available. It makes reaction searching simpler and easier, particularly for the occasional user. It is described in some detail on page 4.

This year's **CDS User Meeting** will be held on July 15th at Daresbury. See page 3 for more details. All users are invited to attend.

Users are reminded that publications stemming from the use of the service should acknowledge such usage by citing the following reference:

*The United Kingdom Chemical Database Service, Fletcher, D.A., McMeeking, R.F., and Parkin, D., J. Chem. Inf. Comput. Sci. (1996), 36, 746-749.*

**ISIS/Client** version 2.1.4 is now available for downloading from the CDS website for PowerMacs and PCs. Version 2.1.3 is also still available and this is the last version which is supported on Silicon Graphics, 68000 based Macs and Windows 3.1. From version 2.1.4 onwards the only supported platforms will be PowerMacs and Windows 95 or NT PCs.

The **solid-phase synthesis** database trial has been completed and, based on a survey of users, we have decided to acquire Synopsys' Solid-Phase Synthesis database. This database is currently still available whilst we arrange a permanent licence.

### Contents

New CDS server .....	2	New features from	
CDS User Forum.....	3	the CCDC.....	10
ISIS Reaction Browser 2..	4	Databases available .....	11
SpecInfo courses .....	9	Useful Information.....	12

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



## New CDS server

The current CDS main server, cds1, is now over 4 years old and ripe for replacement. A new machine is now available and we have started moving users over from cds1. We envisage a 3 month overlap period before the cds1 service is terminated, on 1st September 1998. All users will be required to reregister on the new machine via a web form. Details of how to do this are given at the end of this article.

The new machine is a Silicon Graphics Origin 200 Server running IRIX 6.4. It has two 180MHz R10000 CPUs, 256 MB of main memory and 30 GB of disk. It's raw processing power is an order of magnitude better than that of cds1 whilst the disk I/O rate has doubled. In addition it has a much faster network connection and sits on a higher speed ethernet backbone at Daresbury.

Currently all of the CDS packages are available on the new machine. With the exception of the Brookhaven Protein Databank (PDB), they all operate in the same way as on cds1. The PDB search packages **brook** and **pdbfind** are no longer available; the data can be searched via the quest interface with the **questp** command, or via the web from one of the pdb mirror sites.

Most of the utility programs have already been ported. An exception is **xmol**, which will not run under IRIX 6.4. Users should use **rasmol** or a new program, **viewmol** (which should be available soon), instead.

Moving to a Silicon Graphics system will allow us to trial two new software systems:

MDL's Chemscape - web access to the ISIS databases

SpecInfo/ISIS link - allowing access to the SpecInfo package from ISIS/Base

We hope to arrange trials of both these products this year.

### How to reregister on the new server

All users will need to reregister on the new server. This will allow us to update or verify our user address and email data, simplify future administration of user accounts and avoid moving over accounts which are no longer in use.

Reregistration is via a web form at the following URL:

<http://cds1.dl.ac.uk/cgi-bin/r/sgi/test.html>

There is also a link from the CDS website.

The form will prompt you for your cds1 id and password, then display your contact details, which you should update if necessary. You should then be provided with a password for the new server and details of how to access it. Additional links give information on how to transfer files from cds1 to the new server and what ISIS users need to do to make use of the new server.

Please note that ids on the new server are for individuals not groups. If you have been sharing somebody else's id, please apply for your own id via the online registration form at URL:

<http://cds3.dl.ac.uk/reg/regform.html>.

---

## CDS User Forum

The CDS User Forum is an annual event held at Daresbury. This year's meeting will be on Wednesday July 15th. All users are invited and user reps are particularly encouraged to attend.

It is an important event at which we get to hear the views and opinions of the user community and gives an opportunity to discuss the future direction of the service. It is also a good time to talk to the CDS staff about specific problems.

Specific topics for this year include:

- The new CDS server
- The new reaction browser for ISIS
- New databases
- Trials of new systems

In addition there will be time to discuss any matters raised during the meeting as well to run demonstrations of the various packages.

As in previous years we will be covering the traveling expenses of attendees.

If you would like to attend please contact Dave Fletcher at the address given on the back of this newsletter.

## ISIS Reaction Browser 2

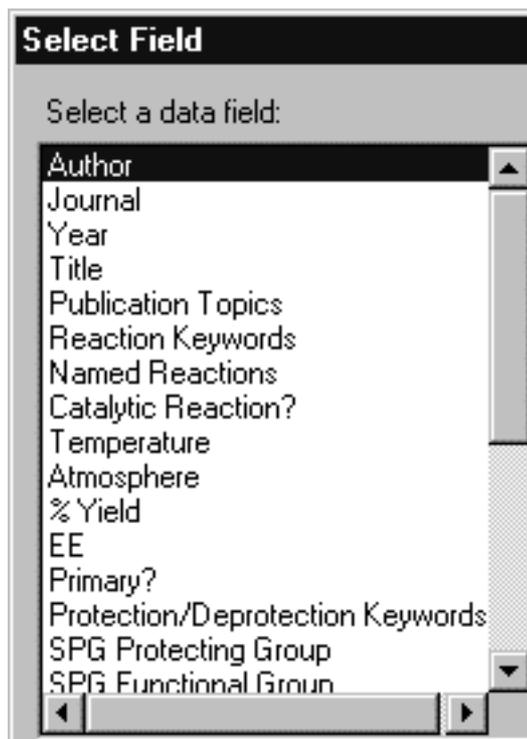
This is a new version of the currently available browser (v 1.3). It has a host of new features and makes database searching easier, particularly for the occasional user. Details of how to download the new browser are given at the end of this article.

### What's new ?

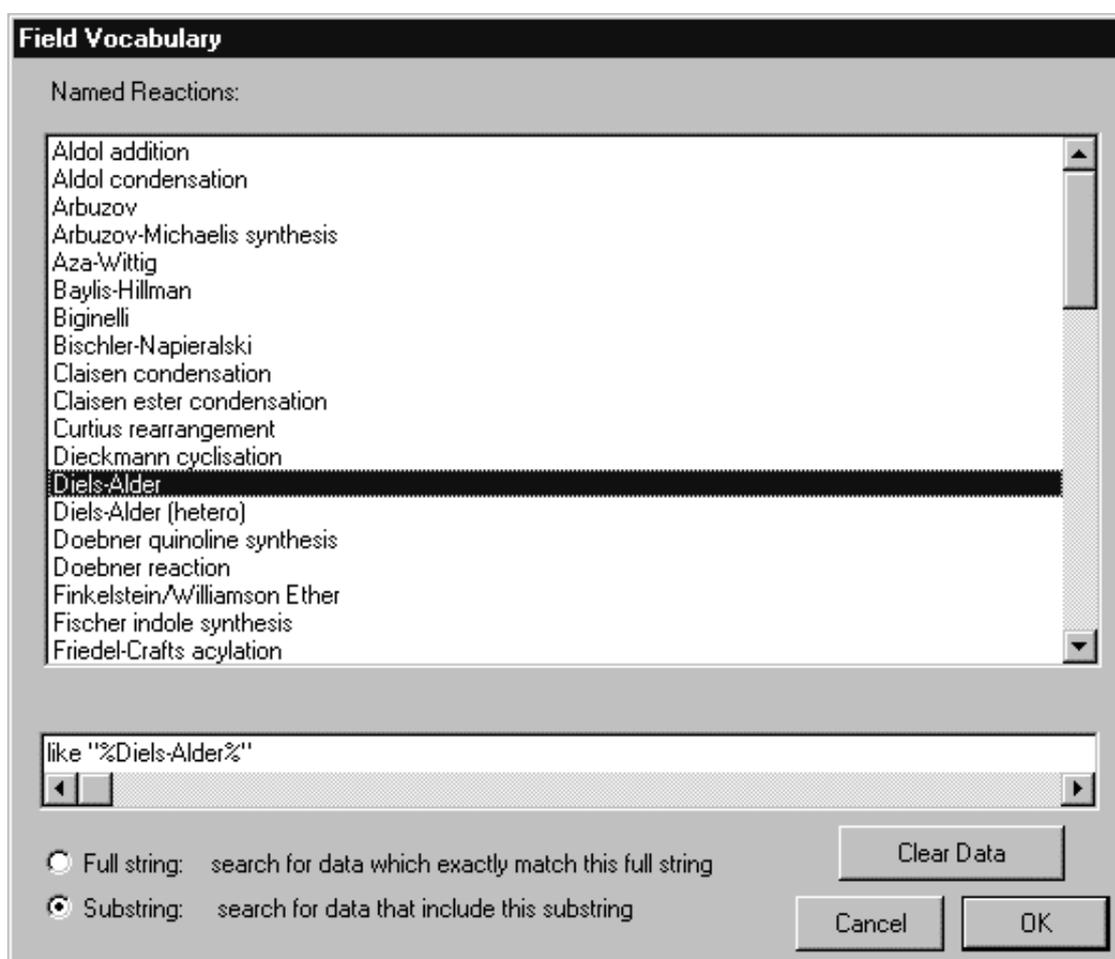
- The browser allows non-MDL databases (such as the Synopsys protecting groups database and solid-phase synthesis database) to be searched and viewed using their own datafields and forms, respectively. Databases can be included or excluded from the search by using the Select Databases... option under the Search menu.
- The query assistant helps build up a query step-by-step for the novice or occasional user.
- Expert query mode by-passes the step-by-step set up of the query and takes you directly to a form to input queries.
- Allows you to easily specify molecule role and type.

Select Role and Type for Molecule Search	
Role of molecule in reaction:	Type of search:
<ul style="list-style-type: none"><li>Solvent</li><li>Reagent/Catalyst</li><li>Reagent/Catalyst or Solvent</li><li>Reagent/Catalyst or Reactant</li><li>Product</li><li>Product not Reactant</li><li>Reactant</li><li>Reactant not Product</li></ul>	<ul style="list-style-type: none"><li>Substructure (SSS)</li><li>Exact Match</li><li>Name</li><li>Formula</li><li>Similarity</li></ul>
Description of the currently selected role of molecule:	Description of the currently selected type of search:
Finds reactions that use the specified molecule as a solvent.	Searches for molecules that contain the specified structural fragment.

- Allows you to easily choose a data field to search from a list which includes data fields for the non-MDL databases SPG and SPS.



- A vocabulary box is then displayed for the chosen field, which helps you select the value to search for.



Vocabulary files are stored on your own desktop machine. The first time a data box is chosen in a session, the browser checks the dates of your vocabulary files with the master files on the host. If newer master files are available, you are given the option to download them.

- Two new types of reaction search are available :

#### Automatic Search

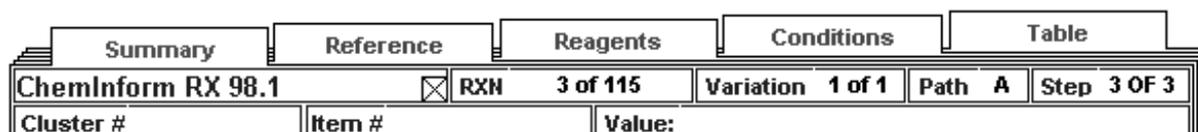
Performs a series of reaction searches, each one with less specificity (Exact Match to Sub-Structure to Similarity), until a relevant example is retrieved from the database(s).

#### Same Transformation

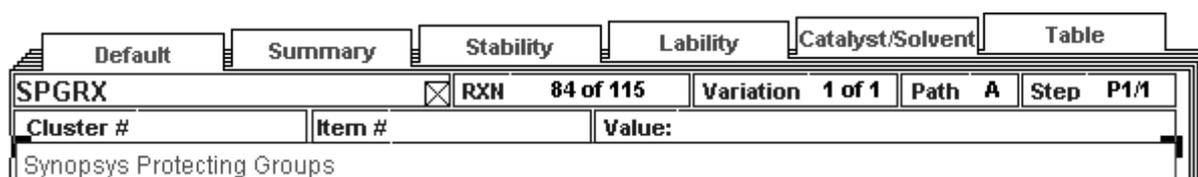
Utilizes the classification codes to generate a reaction query that contains the main reaction transformation, that is, reactions, similar to the one drawn, that differ only in the substituents (R) not involved in the reaction.



- Tabs at the top of the hits display allow you to easily change the display form, as well as showing what data forms are available.



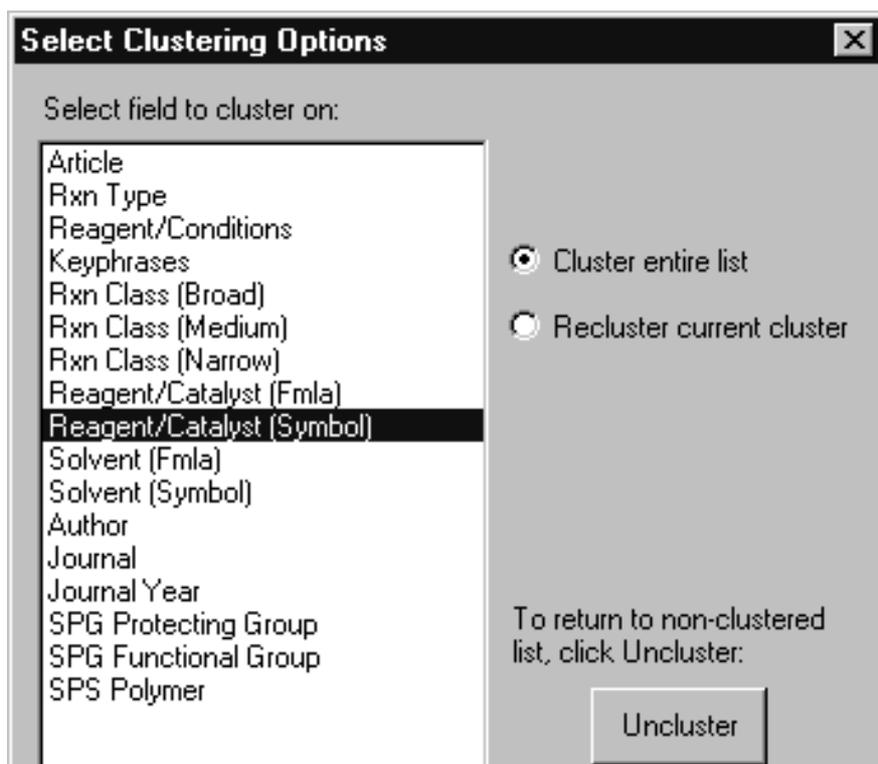
- The browser displays the correct set of forms for each hit, which can be different for each database.



- A number of new options are available on the menu bar.



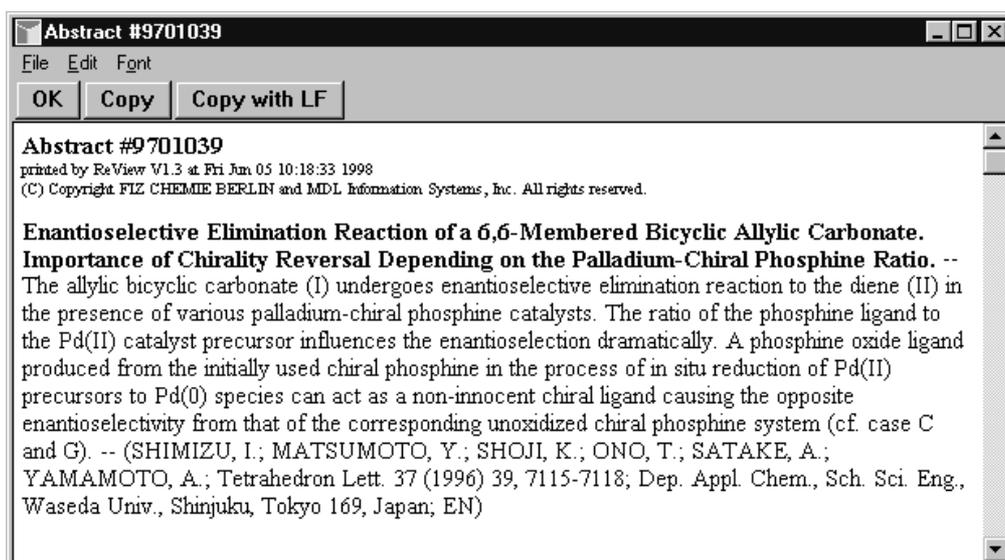
Hits can now be clustered using the **cluster** button on the top menu bar. Clicking this button again reclusters (or unclusters) the current list on a variety of different fields. Clustering is the act of grouping the retrieved reactions into separate clusters depending upon the value of some predefined criteria (such as author or journal).

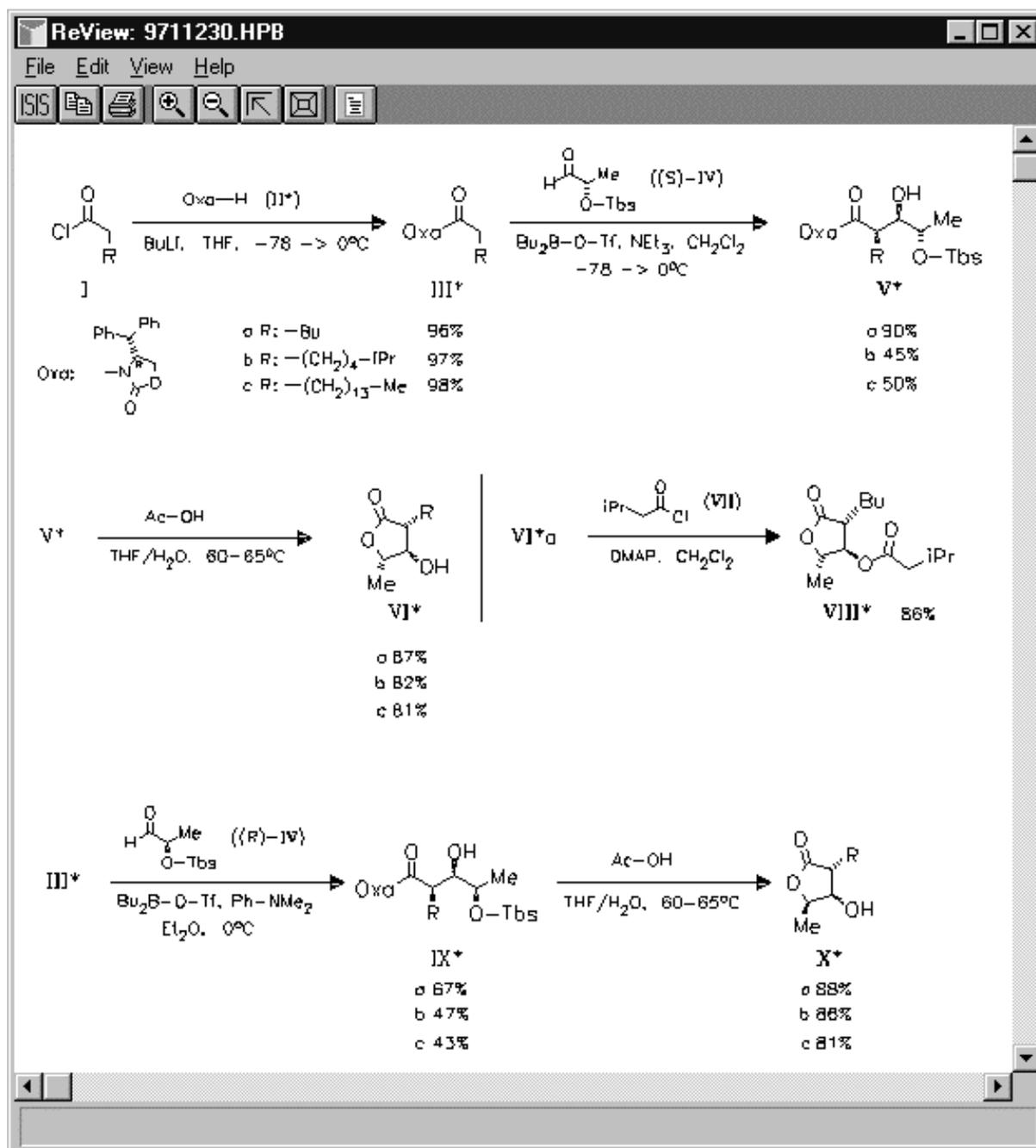


The **examples** button finds other reactions, from databases which are not currently selected, with the same cluster value.

For the ChemInform database only, the **view scheme** button now allows the whole reaction scheme, i.e. the complete synthesis from the reference rather than just the individual reaction, to be displayed (see page 8). For other databases, the button is greyed out and does nothing.

Schemes are displayed by a separate program called review. Various buttons allow resizing, zooming and printing of the scheme as well as displaying the abstract of the paper from which the scheme is taken (on the Mac these functions are accessible via the top menus).





A full reaction scheme displayed using the **view scheme** button from the ISIS Reaction Browser 2 menu bar.

## Downloading ISIS Reaction Browser 2

The browser is only available for the PC and Mac. There is no SGI version. CDS users can download it from the CDS website by following the Information for Users link.

ISIS Reaction Browser 2 requires ISIS/Base 2.1.3 or higher. We have only recently made this version available to users. Again, CDS users can download it from the CDS website.

A brief guide to using the new browser is available at the URL:

<http://www.dl.ac.uk/CDS/browser2/browser2.html>

---

## SpecInfo courses

SpecInfo is a spectroscopic database package designed to aid in the elucidation of structures using spectroscopic data. With the addition of proton NMR and mass spectral data sets in the next few months, the program will become even more effective. We believe SpecInfo is a valuable resource which should be utilised to the full, so this is a reminder that training courses in its use are available.

For beginners and the occasional user we have a half day course on NMR spectrum prediction using SpecInfo. This is the most used feature of SpecInfo and the course covers the following aspects of the program:

- Structure drawing and importing
- NMR spectrum prediction
- Critical examination of the predicted spectrum
- Printing and saving data

A longer course, which also covers the spectral and structure searching features of SpecInfo, is also available. A web based version of the basic course is available from the documentation section of the CDS website.

Courses are offered at no charge to attendees and can be held here at Daresbury or at your Institute (assuming suitable facilities are available).

Courses or workshops can be tailored to fit the needs of particular groups of users; for example, we could run an advanced workshop on NMR spectrum searching or the basic course with an emphasis on particular types of compound.

If you are interested in a course or would like to discuss possible options for a course or workshop, please contact Dave Fletcher at the address given on the back of this newsletter.

## New features from the CCDC

The Cambridge Crystallographic Data Centre (CCDC) have added some new features to their suite of programs:

### **New electronic documentation**

New electronic documentation from the CCDC includes:

- Getting Started with the CSD. This is an enhanced and updated version of the manual that was previously only available in print.
- Volume 3 (appendices) of the CSD user manual
- Guide to new features in Pluto

This documentation is available via links from the CDS website.

### **IsoStar**

IsoStar is a database of experimental and theoretical information on non-bonded interactions. It contains information derived from the Cambridge Structural Database (CSD), the Brookhaven Protein Data Bank (PDB) and molecular orbital calculations. It uses a simple Netscape-Rasmol graphical interface and is available from the CDS via X-Windows.

IsoStar contains information about the geometries, energies and frequencies of occurrence of thousands of different types of non-bonded contacts and is therefore ideal for identifying unusual attractive interactions. Information is presented in the form of scatterplots which show the spatial distribution of non-bonded contacts between two chemical groupings (the "central group" and the "contact group"). Data are stored for hundreds of central groups (terminal substituents, acyclic linking groups, heterocyclic ring systems) interacting with many types of contact group (hydrogen bond acceptors, hydrogen bond donors, hydrophobic groups, ions, etc.). Scatterplots are hyperlinked to the CSD and portions of the PDB, enabling users to inspect individual contacts.

IsoStar requires an X-Windows session to cds1, where it is started by typing "isostar". Online help is available within the application, and more information is available by typing "news isostar" or from the CCDC website at

<http://www.ccdc.cam.ac.uk/prods/isostar.html>

## Summary of 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 777,000 searchable reactions and supplier information on 240,000 different compounds.
- 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 and 20,898 infra-red spectra. It does not, at present, contain any mass spectra.
- CSD** The Cambridge Structural Database. Crystal structure data for 181,000 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 46,000 inorganic structures - the companion file to the Cambridge organic file.
- MDF** Metals Data File. Crystal structure data for nearly 56,000 metals, alloys and intermetallics.
- PDB** The Brookhaven Protein Data Bank containing bibliographic and coordinate details for proteins and other biological macromolecules. There are currently over 7,600 coordinate sets.
- CDIF** Crystal Data Identification File. Crystal class and unit cell data for over 210,000 crystal structures.
- FNMR** A databank of 6,000 <sup>19</sup>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](mailto: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 [cgsbb@dl.ac.uk](mailto:cdsbb@dl.ac.uk)

## 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) 603162

## For QUEST specific problems:

Dr. R. F. McMeeking  
Email [R.F.McMeeking@dl.ac.uk](mailto: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://www.dl.ac.uk/CDS/cds.html>

## 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.

## 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.

## cds1 (main service machine):

Internet number: 148.79.160.113  
Internet name: cds1.dl.ac.uk

## New CDS server

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](mailto:D.A.Fletcher@dl.ac.uk)