



---

# Chemical Database Service

## NEWSLETTER

Issue 7

April 1997

---

The **Beilstein/CrossFire** service, which has been successfully hosted by the CDS, will be moved to the MIDAS centre this spring. We will continue to provide access to Beilstein/CrossFire until July, but users are encouraged to transfer to MIDAS as soon as possible. The database at CDS will not be updated again, so the MIDAS version will contain more up to date data. In addition the MIDAS service will be providing trial access to Beilstein Abstracts. The article on page 2 gives more details as well as instructions for transferring to the MIDAS service.

This year's **User Forum** will be held on July 9th at Daresbury. All users are invited to attend. See page 3 for more details.

A new version of the **ISIS** client software (ISIS/Client 2.1.1) is now available for PCs and Macs. This version for SGIs should be available late summer this year. The software can now be downloaded from the CDS web site (see below) but is also available via ftp from cds1 from the directories /ftp/isis/ibm or /ftp/isis/mac.

A number of new features have been added to the **CDS web** site. These include the ability to **download software** (such as the ISIS/Client software) and documentation. There is also an online database of common problems and frequently asked questions about CDS packages. The article on page 7 gives more details of how to access these new services.

The trial of the **chirbase** chiral separation database is due to finish at the end of June. We would like comments from users who have tried the database in order to decide whether or not to acquire it. Details about the database and how to access it are in the previous CDS newsletter, which can be read online on cds1 via the 'ois' command.

### Contents

Beilstein.....	2	Synopsys Protecting	
CDS User forum .....	3	Groups Database .....	9
SpecInfo .....	4	Crystallography update .	10
ISIS via X-Windows .....	6	Databases available .....	11
New Web services .....	7	Useful information .....	12
Admin commands.....	8		

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



## Beilstein CrossFire Service Transfer

In keeping with JISC (the Joint Information Systems Committee, whose finance underpins the current CrossFire deal) policy the current Beilstein CrossFire service will be transferred from the Chemical Database Service to one of JISC's data centres, MIDAS at the University of Manchester, this spring.

The Beilstein CrossFire Service at MIDAS will run on a dedicated IBM RS/6000 Model 590 server, which is a more powerful machine than the current host, cds2. All users will need individual usernames and passwords to use the service. An online self-registration system is in place to facilitate this. Users will need to get the site access code and password from their local CrossFire technical contact in order to use the self-registration system. A list of local contacts is available on the MIDAS web pages. Note that these local contacts are not the same as departmental CDS reps.

Type "midas" on cds1 for the latest details of the transfer.

### Timescales

The following timescales for the transfer of the Beilstein CrossFire Service from CDS to MIDAS have been decided by JISC after consultation with the various parties concerned.

April 1997	Preliminary documentation and preliminary user registration for CrossFire at MIDAS.
6 May 1997	Start of CrossFire Service at MIDAS.
May-June 1997	Transfer of existing CrossFire users from Daresbury CDS to MIDAS. Publicity and training schedule available from MIDAS.
30 June 1997	User transfer completed.
July 1997	Final month (for contingency purposes only) of CrossFire Service at CDS.

### Abstracts Trial

The Beilstein CrossFire Abstracts will be available via CrossFire at MIDAS for a four-month trial from the beginning of May. CrossFire Abstracts is a companion to the CrossFireplusReactions system, and provides access to titles and abstracts from the organic chemical literature, from 1980 to the present. It is fully integrated with CrossFire and is accessed in the same way via the Beilstein Commander client software.

## Eligibility

CrossFire at MIDAS, including the four-month trial of CrossFire Abstracts, will be available to anyone from the UK HE academic community whose institution has signed the CHEST CrossFireplus site licence agreement. For further details of the agreement, contact your local CHEST representative, or see the details and prices on the WWW at <http://www.chest.ac.uk/datasets/crossfire/contents.html>.

## MIDAS

MIDAS is a national research support service based at the University of Manchester, specialising in on-line provision of strategic research and teaching datasets, software packages, training and large-scale computing resources for the UK academic community. MIDAS is a free service for academia throughout the UK and is funded by JISC, the ESRC and the University of Manchester.

For more information see the MIDAS WWW site at

<http://midas.ac.uk/crossfire/>

---

## CDS User Forum

The CDS User Forum is an annual event held at Daresbury. This year's meeting will be on Wednesday July 9th. All user reps are encouraged to attend, though all users are welcome.

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:

- World-Wide-Web services from CDS
- Hardware upgrade for the CDS servers
- Overview of recent developments
- New databases
- Beilstein/Gmelin

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 travelling expenses of attendees.

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

## SpecInfo

The SpecInfo program is an interface to the spectral database held at Daresbury. It allows users to interrogate the database and displays and prints the results of searches. The current functionality of the program lets you:

- Search database for compounds with spectra (NMR or IR) or fragments of spectra matching the query spectrum.
- Search the database for compounds via (sub-)structure or bibliographic information (such as name or CAS number).
- Draw a structure and predict a NMR spectrum for it.

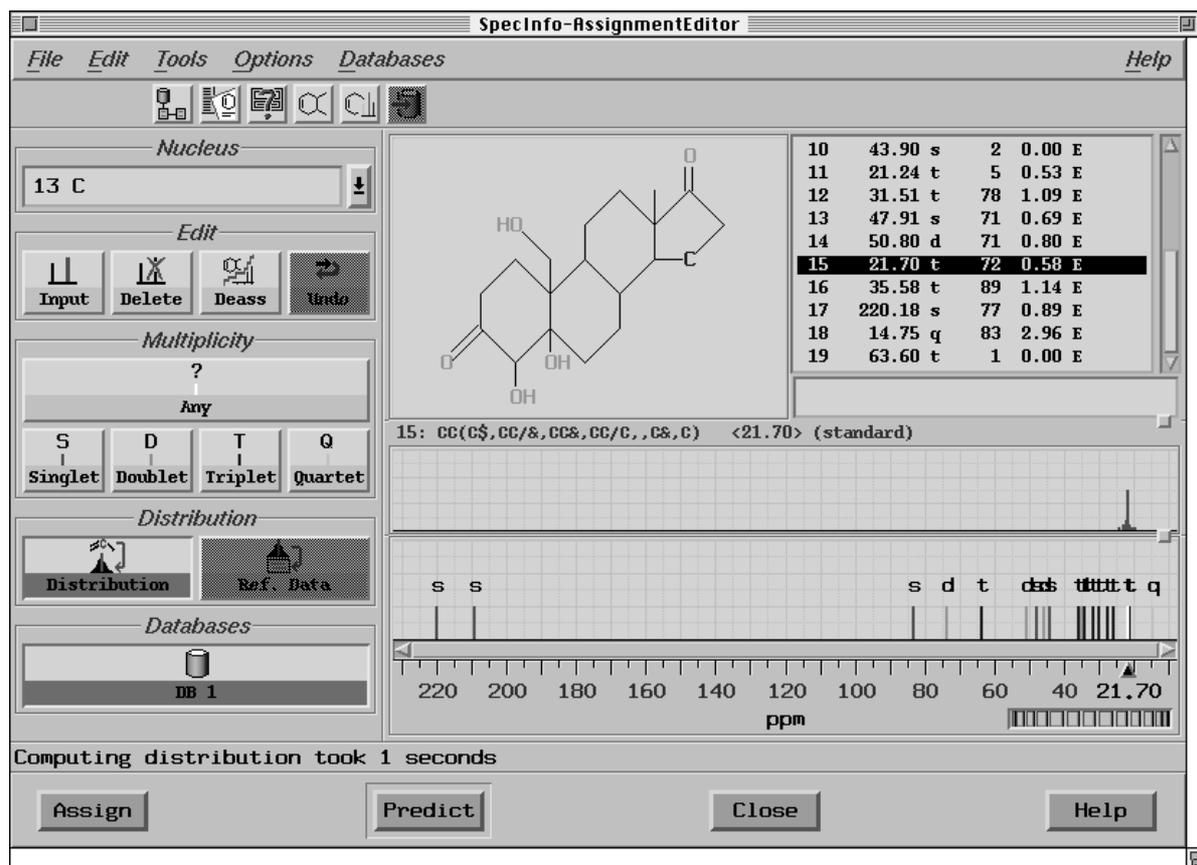
The spectral database at Daresbury currently contains 99,000 C-13 NMR spectra, 6,000 heteroatom (N-15, O-17, F-19 and P-31) NMR spectra and 20,000 IR spectra.

The program was designed as a structure elucidation tool, using a cycle of spectrum searching and spectrum prediction to home in on the structure of the sample from which the trial spectrum was taken.

The most used function of SpecInfo is the NMR prediction. This facility can be used for anything from a simple "look up" of the shifts for a functional group to the simulation of the full spectrum of a complex molecule. The program predicts shifts by a comparison of the chemical environment of atoms in the query structure with those stored in the database. The predicted shift is then the average of all the appropriate shifts from the database. For each shift predicted, SpecInfo gives the number of pieces of data used in the prediction and a standard deviation. In all cases it is possible to examine the original spectra and data from the database used in the prediction.

An example of a predicted spectrum using SpecInfo is shown opposite. The structure was drawn with the SpecInfo structure editor, though it could have been imported from another drawing package, and then the prediction took a few seconds. The output can be saved as a PostScript file for printing.

In this example the predicted shifts are all less than 2 ppm from the spectrum of this compound published in 1990.



SpecInfo can read in spectra in the JCAMP-DX format produced by most spectrometers. For routine analysis you can simply search on the imported spectrum and the program will give you the structures with the closest matching spectra. The database has spectra for nearly 100,000 compounds. If your compound is present it will be top of the hitset. Even if your compound is not present, the structures giving the closest match spectra can give vital clues for the structure of the unknown.

SpecInfo is being continuously developed, with new features being added all the time.

#### Recent new features include:

- A link from the predicted spectra to the reference data used in the shift estimation.
- Proton NMR searching facilities, though currently we do not have any proton NMR data in the database.

#### Features under development, with expected release dates:

- Mass spectra searching facilities and a dataset of 10,000 spectra - Autumn 1997.
- Proton NMR spectrum prediction - Autumn 1997.
- C-13 NMR coupling prediction - 1998.
- IR spectrum prediction.

**Possible new features which could be acquired (assuming sufficient demand from the community and available resources):**

- Proton NMR dataset, 1,000-3,000 spectra should be available Summer 97.
- Link to ISIS, allowing access to SpecInfo via ISIS/Base.
- An additional 50,000 mass spectra dataset.

There is also the facility to add new data to the database, thus improving the quality of the predicted spectra. It would be particularly useful to add data from classes of compounds which are not well represented in the current database.

## ISIS via X-Windows

The organic reaction databases available from CDS are all accessible via the ISIS client/server software. Currently the client software is only available for PCs, Macs and SGI computers. However, we do have a system in place to allow access for X-Windows users who do not have a suitable PC, Mac or SGI to run ISIS/Client.

To use ISIS via X-Windows, you need to do the following:

Log on to cds1 and copy all the files from the directory /ftp/isis/support to your home directory (this only needs to be done once). This can be done with the following command:

```
cp /ftp/isis/support/* $HOME
```

Make sure you have set the DISPLAY environment variable (using `setenv DISPLAY <IP>:0`, where <IP> is your IP or Internet address) and allowed access to your display from cdssg1 (148.79.160.164).

Type `isisbase` to start.

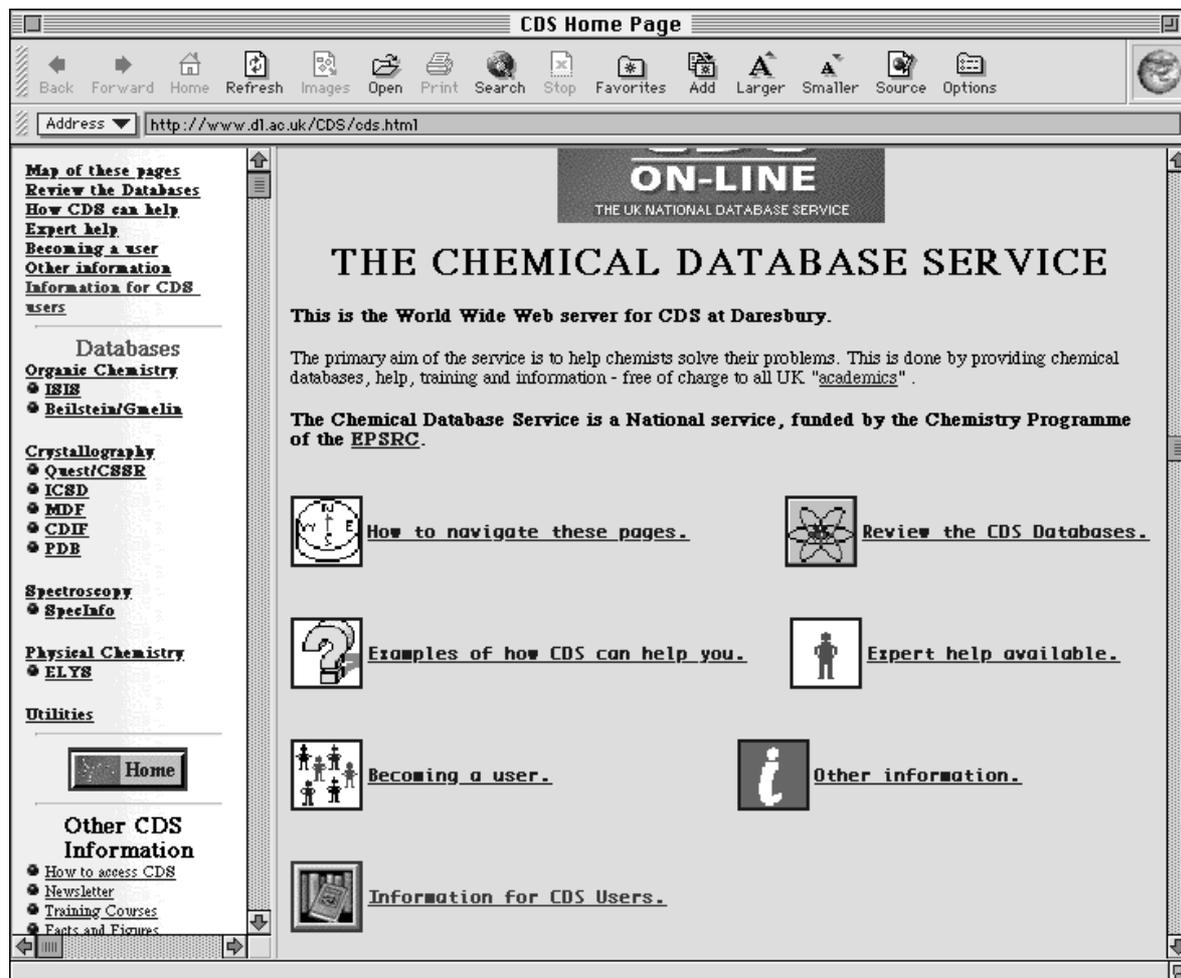
Choose Open from the File menu and select one of the following databases:

all.db	Main reaction databases
acd.db	Available Chemicals Directory
spg.db	Synopsys Protecting Groups Database
chibase.db	Chiral Separation Database (on trial)

Only version 1.2 of ISIS/Base is available at the moment, though we expect an update to version 2.1 soon. In addition this method of using ISIS does not give any of the additional RXL Browser functionality, such as the clustering of hits.

## CDS web site

The CDS web site (URL <http://www.dl.ac.uk/CDS/cds.html>) has been upgraded, and a number of new features added for CDS users. These can be accessed via the new "Information for CDS Users" link on the home page.



New features added for users include:

- Web download area  
Download ISIS and Beilstein client software, CDS manuals, documentation and registration forms over the web.
- Problem and frequently asked questions database  
A small but growing collection of problems and frequently asked questions about CDS packages together with their solutions.
- User representative finder  
Find who is the CDS user representative for your department.

## Information for CDS users

This area contains services and information intended for current users of the Chemical Database Service. Much of the information from the Online Information Server on cds1 will eventually appear here. Some sections on this server are password protected - use your cds1 id and password when prompted.

### Services and information available

[CDS download area](#) - for downloading client software, documentation and forms.

[FAQ/Problem database](#) - solutions to common problems.

[Reps list](#) - find the CDS user representative at your department.

Some sections, such the ISIS/Client software download area, are password protected, to restrict access to CDS users. If you are prompted for an id and password use your usual cds1 username and password.

The online CDS manuals and documentation will be added soon together with news and login messages and an online registration system for new users.

In the longer term it is intended that the web site will completely replace the current Online Information Server (OIS) on cds1. This has the advantage of allowing users access to the OIS information via their own desktop browsers rather than having to run a browser on cds1.

## Admin commands

There are a number of commands on cds1 for checking the status of your account and associated details.

- |               |  |
|---------------|--|
| <b>myinfo</b> | New command to check and update some of your registration details, including your preferred email address. |
| <b>auth</b>   | Command to check which database packages your account is authorised to access.                             |
| <b>quota</b>  | Command to check how much of your disk quota (given in 512 B blocks) has been filled.                      |

## Synopsys Protecting Groups Database

The Synopsys Protecting Groups Database, accessible via ISIS, has been updated. The new release contains 24,333 reactions, an increase of 5%, and 78 new protecting group types. This release covers the primary literature to June 1996. In order to access this new data, users will have to download a new database access file. This is available over the web from the "Information for CDS Users" link on the CDS web site or via ftp from cds1. Ftp users should get one of the following files depending on the type of machine they run ISIS on:

Machine	File
PC	/ftp/isis/ibm/spg.db
Mac	/ftp/isis/mac/spg.hqx
SGI	/ftp/isis/sgi/spg.db

Note that the PC and SGI files are binary.

The Synopsys Protecting Groups Database has been designed to provide chemists with ready access to selected information on the full range of protecting group chemistry. The nature of this chemistry requires an emphasis on reaction conditions and conditions under which the protecting groups are stable or labile. Stability data is available for 61% of the database entries, whilst lability data is available for 85% of entries.

The database is abstracted from the primary literature and updated annually. Protection, deprotection or transprotection reactions are included if they meet one or more of the following conditions:

- the protecting group is new to the database
- the protection step conditions are new to the database
- the deprotection step conditions are new to the database
- the protection or deprotection step demonstrates important regio- or chemoselectivity

Analysis of the database yields the following statistics about functional groups, reaction and substrate types.

### Functional groups:

Alcohols	30.7%	Aldehydes	5.1%
Amines	19.2%	Phosphorus acids	3.6%
Carboxylic acids	17.3%	Thiols	2.7%
Ketones	8.1%	Amides	1.1%
Diols	5.2%	Alkenes	0.9%
Others	6.1%		

**Reaction types:**

Protection steps	46%
Deprotection steps	46%
Transprotection steps	7%
Miscellaneous	1%

**Substrate types:**

Peptides	17.5%
Carbohydrates	10.8%
Nucleosides	10.2%
Steroids	3.4%
beta-Lactams	1.7%
Others	56.4%

## Crystallography update

A number of new or updated utilities have become available in recent months. A summary is given below:

- crad**        There is a new version of the CRAD utility program in place. There has been a certain amount of restructuring of the program code, but users should not detect any significant differences other than tidying of the output format. The old version of the program is still available as ocrad.
- free2xr**     This utility allows users to convert their own sets of cell data, coordinates and symmetry information to .XR file format. It takes data, in a relatively free format, from standard input and writes the .XR file to standard output. The CDS standard SYMOPS pointer, as determined from FNDPNT (see below), is used to define the symmetry information. For details of how to use the utility type "news free2xr".
- free2cam**    As free2xr, except that the output is Cambridge FDAT format.
- free2shx**    As free2xr, except that the output is SHELX format.
- fndpnt**      This is an interactive program which allows you to select a space group on the basis of symbol or number. In addition the various possible setting can be inspected. It provides the SYMOPS pointer required by .XR files, which is used internally by various CDS utilities.
- dmpsgs**     This utility takes the SYMOPS pointer as an input and outputs the space group symbol and symmetry operations.

## 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 700,000 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  $^{13}\text{C}$  NMR; 999  $^{15}\text{N}$  NMR; 856  $^{17}\text{O}$  NMR; 2,183  $^{31}\text{P}$  NMR; 1,825  $^{19}\text{F}$  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 168,000 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.
- ICSD** Inorganic Crystal Structure Data File. About 43,000 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 5,800 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  $^{19}\text{F}$  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 and registration of new accounts:**

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 is available over the World Wide Web at URL:  
<http://www.dl.ac.uk/CDS/cds.html>

## **Online help and information:**

Online help and information, including documentation, is available via the Online Information Server on cds1. Type 'ois' to access it.

## **Documentation:**

Most documentation is available online and some can be downloaded from the CDS web site. Printed manuals can still be ordered through the online orderdoc utility or 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.

## **cds1 (main service machine):**

Internet number: 148.79.160.113  
Internet name: cds1.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)