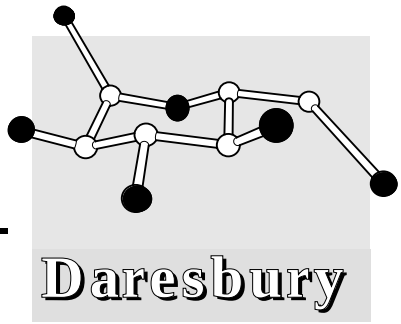
Chemical Database Service

NEWSLETTER

Issue 2 August 1994



Welcome to the second Chemical Database Service Newsletter.

The new service machine, CDS1, is now up and running. As well as the Cambridge Structural Database Software (QUEST, PLUTO and VISTA), the Brookhaven Protein Databank (PDB) and a variety of utilities are now available - see the article on page 9 for more details.

Attached to the back of this newsletter is a questionnaire about the service. It is important for us to find out your views, so we can provide the service you want. Please take the time to fill it in and return it to us at Daresbury.

Our training program in the major packages we provide is now well under way, with courses being given both here at Daresbury and at other university sites. The article on page 10 describes the computing requirements for us to be able to give a course at your site.

A brief reminder that results of research using the Chemical Database Service which are published should include the following acknowledgement:

We wish to acknowledge the use of the EPSRC's Chemical Database Service at Daresbury

There is now a new reference that should be used for acknowledging use of the Cambridge Structural Database (either via CSSR or QUEST).

Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M., & Watson, D.G. (1991) *J. Chem. Inf. Comp. Sci.* **31**, 187-204.

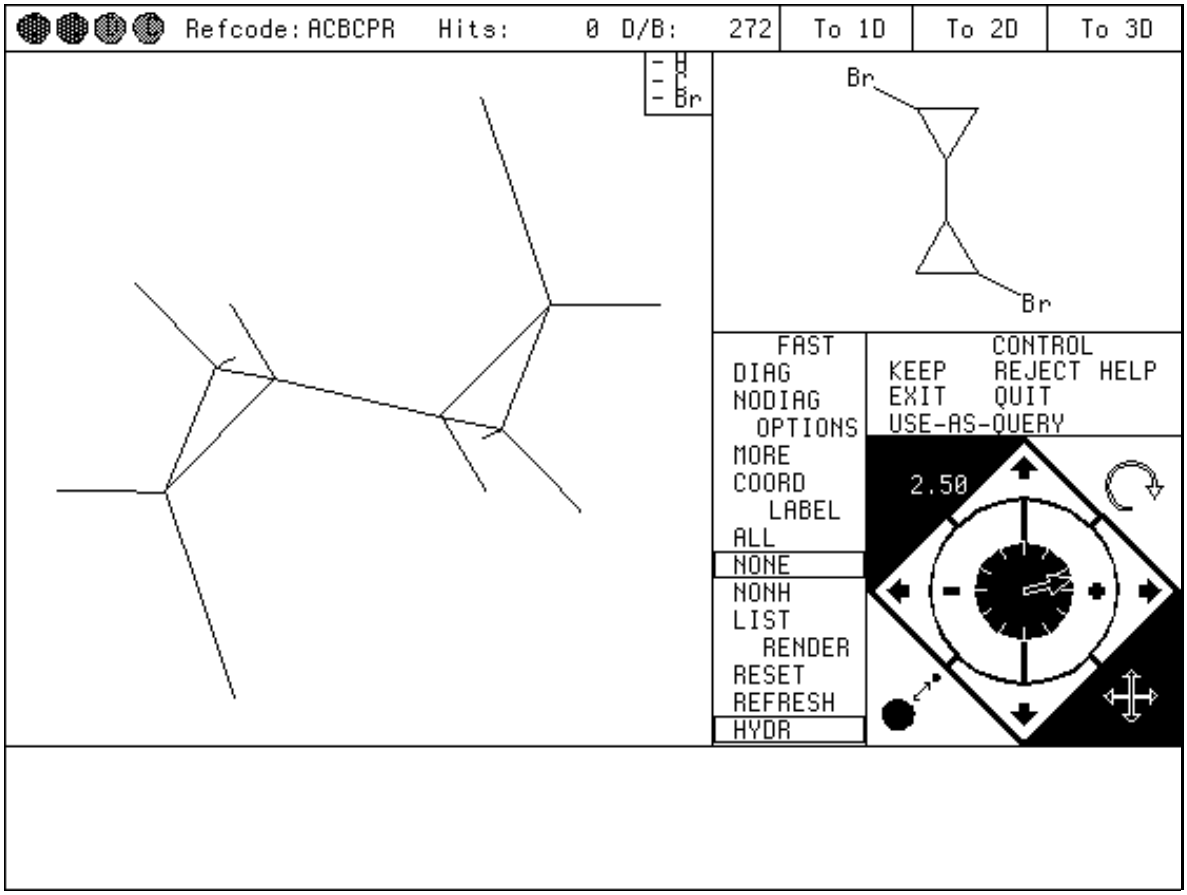
Contents

Cambridge Structural Database System

The Cambridge Structural Database System (CSDS) has been available on CDS1 since April this year. It consists of three separate programs to interrogate and display information from the Cambridge Structural Database (CSD). The programs are currently available to all users. If you are not already registered to use CSSR then please request a registration form from UIG at the address given on the back page.

QUEST is the database search program. It is an interactive search program that uses menu-driven graphics for both query construction and the display of results. It allows the integrated searching of bibliographic and chemical text, together with 2D structural and sub-structural searches subject to 3D geometric constraints.

References can be displayed in a variety of styles: bibliographic and chemical text; 2D chemical structure; 3D, rotatable, crystal and molecular structure or a combined 2D/3D display. A typical reference is shown in the 2D/3D display below.



PLUTO is a program which is used to display database entries previously selected by QUEST. It allows easy visualisation of molecules in 3D in a variety of styles, and also exploration of its environment via H-bonded networks and packing diagrams. PLUTO takes its input from a file created by QUEST. It will also take FDAT files produced by the CDS Utility program, GETFDAT. PLUTO can also read

non-CSD data in an easy-to-type format. The functions of PLUTO will be incorporated into the 3D display in QUEST in a future release of the CSDS.

The purpose of VISTA (Visualisation of STAtistics) is to provide an easy-to-use interactive facility for analysis of statistics for the results of a geometric search of the database. The QUEST program provides basic display of histograms and scattergrams and saves a table of the selected geometric parameters for later analysis. VISTA provides facilities for exploring the data set, checking on the actual data entry in normal 1D/2D/3D displays associated with any selected table entry, and the option to exclude entries from analysis if required. Ranges of parameters can be selected, and there are extensive Postscript editing features for output of histograms and scattergrams for printing.

This is the first release of the program which provides the basic analysis functions for histograms, scattergrams, regression and PCA. It will be further developed to include all of functions of GSTAT, which will continue to be available as a non-interactive program.

The CSDS Getting Started Manual has been sent out to all CSSR users and will be provided to all new users. A fuller set of documentation is available directly from the Cambridge Crystallographic Data Centre. In the future we hope to have the full set of manuals available online.

The Brookhaven Protein Databank files are now accessible through the CSDS. Bibliographic and sequence information may be searched with QUEST. In the future it is possible that the 3D coordinate information will also be included. To search just the Brookhaven files use the command *questp* in place of *quest*. To search the CSD and Brookhaven files use *questa* in place of *quest*.

Recommended Terminal Emulators

REACCS, SpecInfo and QUEST all require a graphical terminal emulator in order to display output correctly. We recommend the following emulators. We have tried all of them and therefore are able to give advice to users.

For IBM PC	Emu-Tek
	Vista eXceed (X-windows emulator)
For Apple Mac	VersaTerm Pro
	MacX (X-windows emulator)

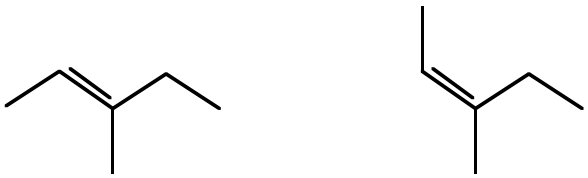
Both IBM PC products are available through CHEST.

The two X-windows emulators have some problems with some of the CDS packages. See the article on X-windows access in this newsletter for more details.

Stereochemistry in SpecInfo

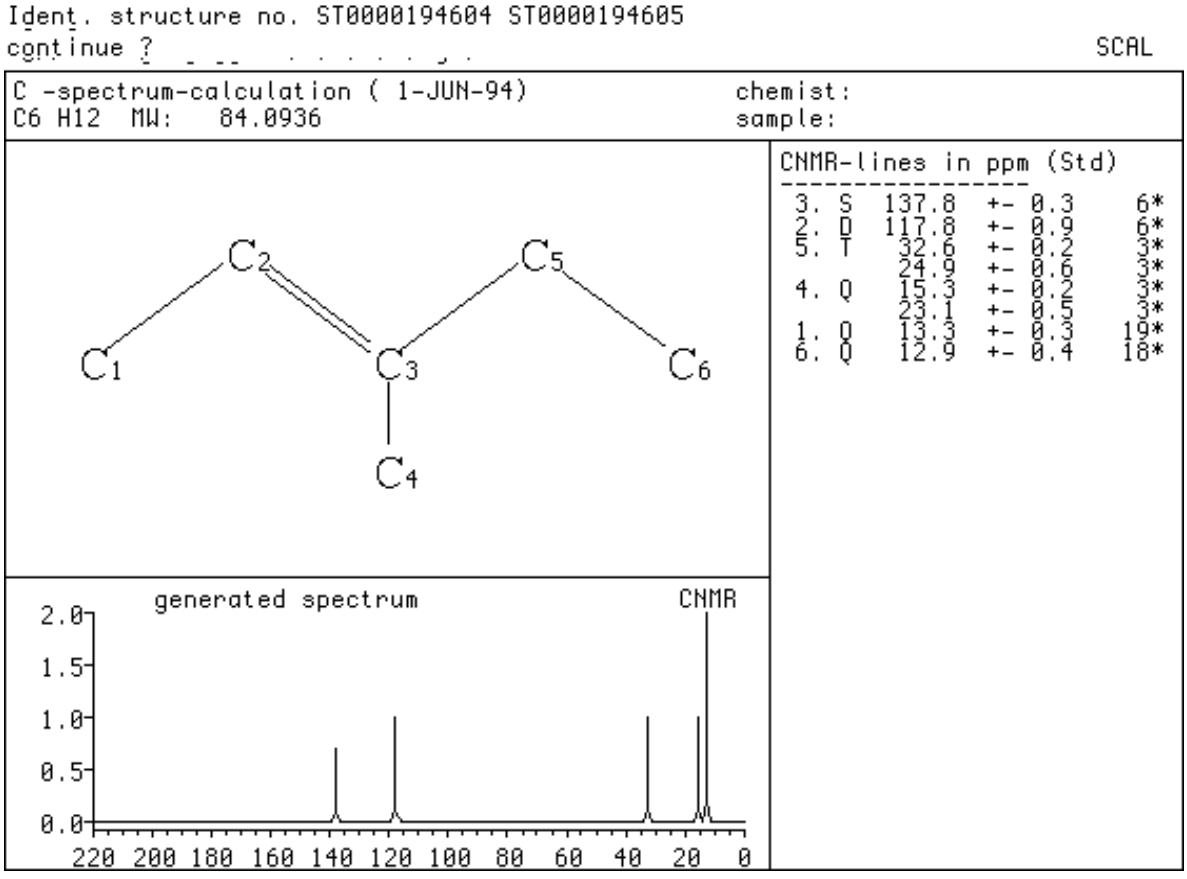
SpecInfo is a multi-technique spectroscopic database package covering NMR and IR spectroscopy. The program is designed to aid the chemist in spectral interpretation and structure elucidation problems.

Structures are stored within SpecInfo in terms of atoms and the connections between them. Such connectivity information, however, takes no account of the 3 dimensional nature of the structure and thus cannot distinguish between different geometric or stereo isomers. As far as SpecInfo is concerned the following two structures are the same.



Obviously, the ^{13}C NMR of these two isomers will be different. In order to be useful, the program needs to be able to deal with this situation correctly when it calculates a spectrum. So how does SpecInfo handle this?

A spectrum calculation using either of the above structures will produce the following:



Note that for carbon atoms 4 and 5 the program has produced two calculated shifts - one for each isomer. What SpecInfo has done is to detect a double peaked distribution in the stored shifts for each chemical environment. Rather than just averaging them all together, it averages each distribution separately. Thus it successfully deals with the problem without any 'knowledge' of stereochemistry. It is up to the chemist to determine which shift goes with which isomer, either by chemical knowledge or by tracing the calculation back to its references (using the Sn command, where n is the number of the atom in question).

CDS User Forum

The first Chemical Database Service User Forum was held on May 17th at Daresbury Laboratory. It was chaired by Dr. Marjorie Harding from Liverpool University and attended by 27 user representatives from sites across the UK. The meeting had two main objectives:

a) to make user representatives aware of the wide range of services available and to enable them to help others make use of these

b) to get feedback from the user community and constructive comments on all aspects of the service.

The meeting was a success. The current service and future changes were discussed and a variety of user queries were addressed. It emphasised the key role that Chemical Database Service User Reps can play in facilitating use by other researchers in their Departments.

It was agreed that there should be future meetings approximately once a year.

Minutes and details of the meeting are available from Don Parkin at the address given on the back of this newsletter.

Summary of databases available from the Chemical Database Service

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

REACCS A chemical reaction information management system allowing search, retrieval and display of molecules, reactions and their associated data. Currently contains 239,620 searchable reactions and supplier information on 131,723 different compounds.

SPEC SpecInfo is a multi-technique spectroscopic database package covering NMR, IR and mass spectra. The package is designed to aid the chemist in spectral interpretation and structure elucidation problems. The database currently contains 66,063 ^{13}C NMR; 1,000 ^{15}N NMR; 856 ^{17}O NMR; 2,182 ^{31}P NMR; 1,830 ^{19}F NMR and 28,464 infra-red spectra. It does not, at present, contain any mass spectra.

CSSR/QUEST The Cambridge Structural Database. Crystal structure data for 120,480 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 37,006 inorganic structures - the companion file to the Cambridge organic file.

MDF Metals Data File. Crystal structure data for about 52,500 metals, alloys and intermetallics.

PDB The Brookhaven Protein Data Bank containing bibliographic and coordinate details for proteins and other biological macromolecules. There are currently nearly 2,000 coordinate sets.

CDIF Crystal Data Identification File. Crystal class and unit cell data for 197,612 crystal structures.

FNMR A databank of about 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 11,000 entries.

X-Windows access to the Chemical Database Service

The graphical based programs (REACCS, SpecInfo and QUEST) can all be accessed via X-terminals or emulators. Details of how to start up each program are given below for the emulators Vista eXceed (for IBM PC) and MacX (for Apple Mac) and for a UNIX workstation.

In each case <IP address> is your IP or Internet address, which is a four number code such as 148.79.160.7 . Items to be typed in are given in **bold** text. Note that for QUEST on CDS1, the text is case sensitive.

Vista eXceed:

REACCS
telnet onto DLVB
XIP <IP address>
RUN X_DIR:XTERM
click in the xterm window
REACCS
use terminal type 21c or 21 for monochrome

SpecInfo
telnet onto DLVB
XIP <IP address>
CREATE/TERM
click in the DEC window
SPEC
use terminal type 13

This is not a very satisfactory emulation, since the screen needs to be refreshed manually.

QUEST
telnet onto CDS1
setenv DISPLAY <IP address>:0
quest problemname
term x
menu

MacX:

REACCS
logon to DLVB with the remote command
@CDS$COM_DIR:SXTERM <IP address>
REACCS
use terminal type 21c or 21 for monochrome

This is not a very satisfactory emulation with a small screen, since the REACCS window cannot be resized.

SpecInfo logon to DLVB with remote command
@CDS$COM_DIR:SXTERM <IP address>
CREATE/TERM
click in the DEC window
SPEC
use terminal type 13

QUEST logon to CDS1 with remote command
setenv DISPLAY <IP address>:0;xterm -ls
quest problemname
term x
menu

For MacX access to CDS1, you should add the following line

more /etc/motd

to your .login file, otherwise the login messages will not be displayed.

UNIX Workstations:

There are a number of differences between the various types of workstations. Suns need the xhost+ command in order to allow displays from outside machines. Silicon Graphics Indigos and all but the most recent Indys do not display the text correctly with SpecInfo.

REACCS logon to DLVB
XIP <IP address>
REACCS
use terminal type 21c or 21 for monochrome

SpecInfo logon to DLVB
XIP <IP address>
CREATE/TERM
click in the DEC window
SPEC
use terminal type 13

QUEST logon to CDS1
setenv DISPLAY <IP address>:0
quest problemname
term x
menu

New UNIX machine - CDS1

CDS1 is the name of the Chemical Database Service's new IBM RS/6000 UNIX machine. This machine is set to become the main machine for the service in the near future. All CDS users have been given accounts on the new machine and should have been sent passwords. If you have not received a password, please contact the Computer Help Desk/UIG at the address given on the back of this newsletter.

The following packages are currently available on CDS1

- **Cambridge Structural Database System** (QUEST, PLUTO and VISTA). See the CSDS article in this newsletter.
- **PDB** - Brookhaven Protein Databank.
- **Rasmol** - an X-windows 3D molecular structure viewer. This can be used to display Brookhaven Protein Databank (PDB) files.
- **Xmol** - an alternate X-windows structure viewer, more useful for small molecules
- **Babel** - a chemical structure file format conversion utility.
- **Gopher and xgopher** - ASCII and X-windows gopher clients.
- **Xmosaic** - an X-windows World-Wide-Web client.

We expect to have the following packages available in the near future

- **CSSR** and the other locally written code (ICSD, CDIF, MDF and ELYS).
- **ISIS** - for accessing the REACCS databases.
- **SpecInfo version 3** - a new X-windows version of the program.

Once this has occurred, CDS1 will replace DLVB as the main service machine. DLVB will be retained until at least April 97 to run VAX specific software for those who are unable to access the equivalent packages on CDS1 (REACCS and the old version of SpecInfo).

All users should have received a Guide to Packages on CDS1, which describes how to access the various programs available. A second manual a Guide to UNIX on CDS1 should be available shortly and will be sent out to all CDS1 users.

The preferred method of access to CDS1 is via the Internet. The Internet address is 148.79.160.113, mnemonic cds1.dl.ac.uk. Limited X25 or PAD access is available for users who are unable to connect to the Internet. The NRS name is UK.AC.DARESBURY.CDS1 and the DTE number 00000100026012.

Training

Training sessions for the various database packages are held at Daresbury as well as at outside sites. Details of forthcoming courses are available online on both of the service's machines. On DLVB type COURSES and on CDS1 type news courses for details.

Courses are free of charge for participants, but we are unable to cover travel and substance expenses for those attending. If you would like a course to be held in your area, please get in touch with us at the address given on the back of this newsletter.

To give a course somewhere other than at Daresbury, we need a suitable venue. The basic requirement is a room with 6 terminals capable of accessing the CDS machines and supporting the required database package and an overhead projector. The requirements for the terminals will vary slightly, depending on what database package is being covered, but for REACCS, SpecInfo and QUEST they need the appropriate graphics capability. The requirements for these packages are given below:

Terminals: Tektronix 4105
DEC VT240 or VT340
X-Terminal or workstation

Emulators:

IBM-PC: Emu-Tek 4200 Plus
Teemtalk
Chemtalk Plus
Vista eXceed

Mac: VersaTerm Pro
MacX

Other emulator packages will work, such as Reflection 4 Plus for the PC and eXodus for the Mac, though we have not tested them at Daresbury.

SpecInfo does not work very well under X-Windows, specifically the interface with Vista eXceed is poor and it will not display text properly on most Silicon Graphics machines.

QUEST will also work with Tektronix 4010 emulators, but not with DEC VT240 and VT340.

The Computer Help Desk

On 1st June 1994 the User Interface Group (UIG) commenced running a fully automated Computer Help Desk service at Daresbury. This article will concentrate on how this system will help the CDS user community.

When to use the Computer Help Desk

The first contact point for all CDS general and registration enquiries is still through the Computer Help Desk / UIG. Database and specific technical inquiries should still be directed to the CDS team.

Advantages of an automated Computer Help Desk

- facilitate problem control and management with automated escalation procedures so that all queries will be dealt with promptly.
- provide statistics such as start and completion time of each query to the CDS service manager for comparison against agreed service levels.
- highlight weakly supported and problem areas and well supported and trouble free ones within CDS .
- highlight areas where training of CDS users and DL staff may be required.
- build up an experience or solutions database allowing CDS users and non-specialist DL staff to solve difficult problems themselves.
- improve productivity by ensuring that staff do not have to solve the same CDS related problem twice.

Overall there should be a better and faster response to queries.

How it works

Users don't have to do anything new at all when they have a problem or query. The problem is logged by the front-line staff receiving the call and assigned to appropriate support staff. Users receive automatic email confirmation of when their request is started and when it is finished.

Some statistics

The logging of each incident allows us to produce statistics. Here are some initial very general findings for the first complete month of operation. In June, 81 CDS related queries came in via the Computer Help Desk and at the time of writing (22 / 7 / 94) 79 had been marked closed. This is above the average for all queries. Just over 50% had been closed within 2 hours and 83% were solved by the front-line staff member.

Information

General queries and registration of new accounts:

Computer Help Desk / User Interface Group (UIG)
Room C12, Daresbury Laboratory
Daresbury, Warrington
WA4 4AD
(0925) 603351
Email uig@daresbury.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
(0925) 603162
Email cdsbb@daresbury.ac.uk

For REACCS specific problems:

Dr. D. Parkin
Email
D.Parkin@daresbury.ac.uk
(0925) 603162

For SpecInfo specific problems:

Dr. D.A. Fletcher
Email
D.A.Fletcher@daresbury.ac.uk
(0925) 603162

For QUEST specific problems:

Dr. R. F. McMeeking
Email
R.F.McMeeking@daresbury.ac.uk
(0925) 603162

Documentation:

Documentation is available through the ORDERDOC utility on DLVB or from UIG at the above address.

Access:

The service is available free of charge to UK academic research groups through Janet, Internet, PSS and dialup. Individual user names are issued on request.

DLVB:

DTE address: 000001003000
NRS name:
UK.AC.DARESBURY.DLVB
Internet number: 148.79.160.7
Internet name: dlvb.dl.ac.uk

CDS1:

DTE address: 00000100026012
NRS name:
UK.AC.DARESBURY.CDS1
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
Chemical Database Service (CDS)
Room C18, Daresbury Laboratory
Daresbury
Warrington
WA4 4AD
(0925) 603162

Email
D.A.Fletcher@daresbury.ac.uk