

---

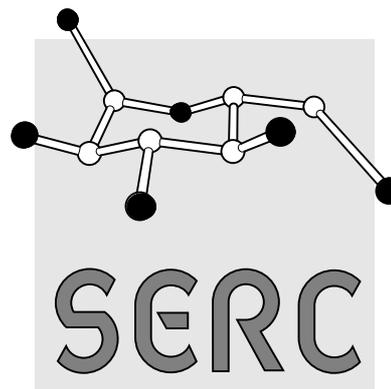
# Chemical Database Service

## NEWSLETTER

Issue 1

February 1994

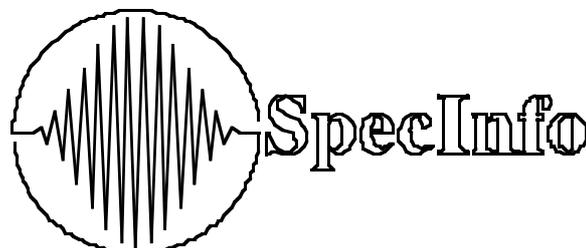
---



### Contents

SpecInfo .....	2	CSD on CDS .....	10
REACCS.....	5	User forum .....	11
Available databases .....	8	Useful information .....	12
Terminal emulators.....	9		
Training courses .....	10		

Welcome to the first issue of the new, improved CDS newsletter from the new improved Chemical Database Service! The service has been expanded with the addition of two new staff and two new systems (SpecInfo and REACCS, described later in this newsletter). This increase in personnel will allow us to provide training courses for the various packages - see the article later in this newsletter.

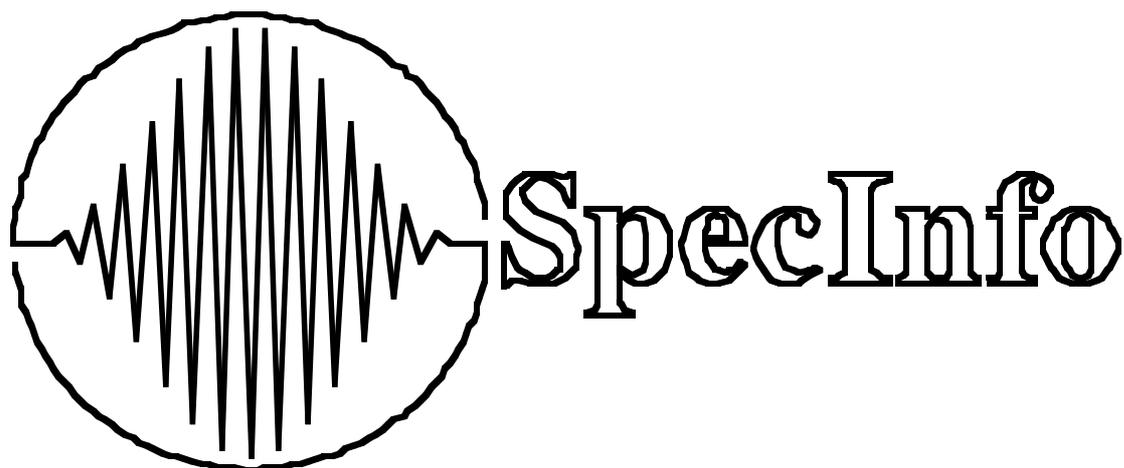


**New packages available from the Chemical Database Service**

Another change in the pipeline is the acquisition of a new IBM RS/6000 computer. This is substantially more powerful than the current  $\mu$ VAX and should be installed sometime in the next few months. PDB and the Daresbury written software (CSSR and others) will be transferred first, followed by SpecInfo towards the end of the year. The REACCS databases will be accessible on the IBM through ISIS. In the long run ISIS will be the only way of accessing the REACCS databases, but we hope to keep both systems running concurrently for as long as is practical. There are no immediate plans to phase out the  $\mu$ VAX.

On the 17th of May this year there will be a CDS user meeting at Daresbury to give users a chance to hear about the service and to discuss the formation of a user group. All CDS users are invited to attend, see page 11 for more details.

In future issues of this newsletter we would like articles from our user community about how they use the service in their research. If you are interested in contributing, please contact Dave Fletcher at the address given on the back page.



## Overview

SpecInfo is a database management system from Chemical Concepts designed to store, retrieve and manipulate NMR, IR and mass spectra. It has been available at the CDS since October 1993. Existing CDS users need to contact UIG in order to have their account set up to run SpecInfo.

The database at Daresbury currently contains 66,000  $^{13}\text{C}$  NMR; 1,000  $^{15}\text{N}$  NMR; 850  $^{17}\text{O}$  NMR; 2,200  $^{31}\text{P}$  NMR; 1,800  $^{19}\text{F}$  NMR and 28,500 infra-red spectra. It does not, at present, contain any mass spectra.

A variety of features are available within the program to aid in the assignment and interpretation of spectra. These include:

- **Spectrum calculation**  
Using the data available in the database, SpecInfo calculates a predicted spectrum for any trial structure. This prediction does not involve any quantum mechanics and its accuracy is obviously dependant upon both the size and diversity of the available database.
- **Spectral search**  
The program will search the database for an exact, or the closest match to a trial spectrum. In addition, fragments of spectra, down to an individual line, can be searched for in the database.
- **Substructure search**  
A variety of structure and substructure searches are available to probe the contents of the database.
- **Hitlist merging**  
The results of different types of searches can be logically combined to produce a powerful multi-technique search system.

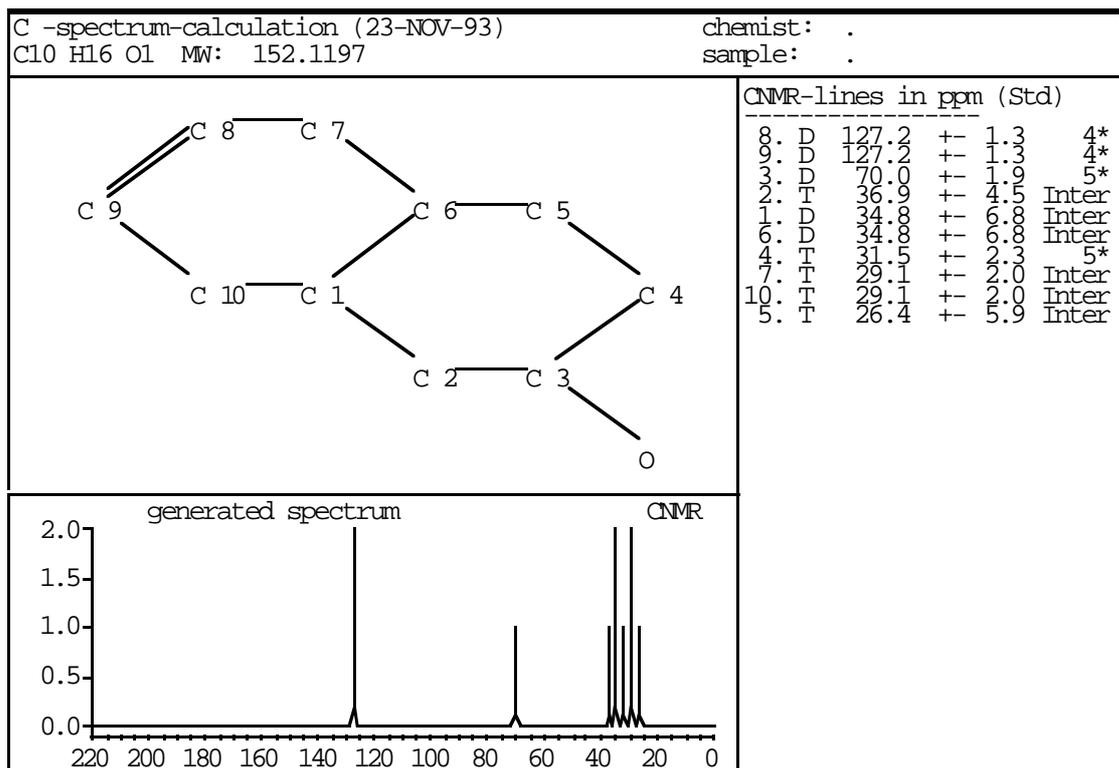
The program also contains a structure editor, a variety of spectra input/output features and an online help facility. In addition it supports the MDL Molfile format for the input and output of molecular structures.

# Spectrum calculation

A typical  $^{13}\text{C}$  NMR spectrum generated by SpecInfo is shown below. The output gives the molecular structure, a listing of the predicted lines together with a graphical representation of the spectrum. The listing gives the multiplicity of the line, its expected chemical shift (with standard deviation) and the number of pieces of data used by the program in the prediction.

continue ? SCALY  
continue ?

SCAL



SpecInfo manipulates structures by using a code for each atom in the molecule. These codes, known as HOSE codes, list the chemical environment for an atom in terms of shells of neighbours. The first shell includes all the atoms adjacent to the atom of interest, the second shell is those one removed, and so on until all the atoms in the molecule have been listed. When the program predicts a spectrum, it takes each atom of the trial structure and generates its HOSE code. The code is truncated at the fourth shell - atoms with the same code for the first four shells are considered by the program to have the same chemical environment. SpecInfo then searches the database for all occurrences of this truncated code and averages the chemical shifts of all the examples it finds to produce the predicted shift. This process is repeated for each atom of the structure to produce the predicted spectrum.

In the example above, SpecInfo predicts a chemical shift of 127.2 plus or minus 1.3 ppm for carbon atom number 8. It predicts the same shift for carbon atom number 9, since as far as it is concerned both atoms have the same chemical environment (i.e. surrounded by the same atoms for four shells).

## Documentation

There are currently three manuals available for the program:

- **SpecInfo Tutorial**  
This is a tutorial introduction to the major features of the program, similar to the tutorial guide available for REACCS. It is recommended for newcomers to SpecInfo.
- **SpecInfo Quick Reference Guide**  
A pocket size summary of commonly used commands with a short explanation of each. It provides a quick reminder of the commands available in any given system and their functions.
- **SpecInfo Reference Guide**  
An almost complete guide to the program provided by Chemical Concepts, the suppliers of SpecInfo. Unfortunately it is only a preliminary version and is quite difficult to follow.

The first two manuals are sent out to everyone who registers for SpecInfo. The Reference Guide, as well as being somewhat unfriendly, is bulky and expensive for us to reproduce so we do not send it out automatically. It can be obtained, free of charge, through the online document ordering service, ORDERDOC, but check whether there is a copy readily available in your department beforehand.

## Online help

Online information and help about SpecInfo is available on DLVB. Simply type S\$HELP at the \$ prompt to enter a VMS style help facility. If you are unfamiliar with this type of help, type README for more detailed instructions.

The file currently contains general information about the program, details of terminal setups, information on printing and some hints on using the program. In addition it includes information about program updates and bugs together with details of CDS SpecInfo training courses.

There is additional online help available within SpecInfo itself.

---

# REACCS

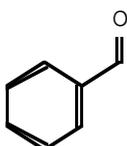
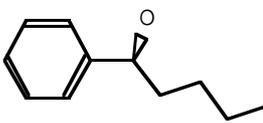
REACCS (REAction ACCess System) is a chemical reaction database management system from MDL Information Systems that is designed to store, search, retrieve, and display molecules and reactions along with their associated data. A variety of chemical databases can be searched within REACCS covering over 50 years of published literature, providing established and current awareness in synthetic organic chemistry, and containing over 210,000 searchable reactions. The databases currently available are:-

- THEILHEIMER (Synthetic Methods of Organic Chemistry, 1946-1980)
- CHC (Comprehensive Heterocyclic Chemistry, compendium up to 1984)
- CLF (The Current Literature File, 1983-1991)
- CHIRAS (Asymmetric Synthesis, 1975-1991)
- METALYSIS (Transition Metal-mediated Methods, 1974-1991)
- JSM (Journal of Synthetic Methods, 1980-present)
- ORGSYN (Organic Synthesis, 1921-1990)
- CSM (Current Synthetic Methodology, 1992-present)

In addition REACCS also allows access to ACDRX (Available Chemicals Directory), which is a database of suppliers of chemicals world-wide. This is periodically updated and currently contains about 129,000 different compounds from 220 suppliers.

REACCS has been available as a component of the CDS suite of databases since June 1993.

Search queries can be based on :-  
 Molecular Structure  
 Atom and Bond transformations in a reaction  
 Reactant, Product, Solvent or Catalyst  
 Reaction Data such as Temperature or Yield  
 Literature References  
 Keyword descriptions of reaction types

Help	Set	Exit	Main	Build	Search	ViewL	Plot	Forms	THEIL:
A + B -> C			First	Next	Prev	Ref=List	ClrRef	Db	Current 274 R
			Item	List	Table	ReadList	WriteList		46785 R
			Data	Omit	Query	DeleteL	IndexList	List	462 R
ITEM 2 OF L100									
<div style="display: flex; align-items: center; justify-content: space-around;"> <div style="text-align: center;">  </div> <div style="text-align: center;">+</div> <div style="text-align: center;">  </div> <div style="text-align: center;"> <math>\xrightarrow[\text{78.0 C}]{\text{1.5 hrs}}</math> </div> <div style="text-align: center;">  </div> </div> <p style="text-align: center;">(2S,2'S)-2-Hydroxymethyl-1-(1-methylpyrrolidin-2-yl)-@ CHIRAL</p> <p style="text-align: center;">77.0% yield</p>									
T Mukaiyama, et al, J Amer Chem Soc, 101, p.1455, 1979									
95% optical yield. For further examples, also with dialkylmagnesiums, see litref(1).									
Theil regno: 274			Extreg: 35439		Path A		1 STEP		Variation: 1 OF 1

A typical reaction output screen in REACCS is shown above.

## REACCS menus

REACCS commands are organised into six main menus which are always displayed in a menu bar at the top of the screen. In addition there is a Help Menu and a Set Menu.

### MENU

### FUNCTION

Main Menu: Add and remove information in a database.  
 Retrieve single molecules or reactions from a database.  
 Create and retrieve external files (molecules or reactions).  
 Choose another database.

Build Menu:	Draw chemical structures and build reactions. Add query features and highlight reaction centres.
Search Menu:	Search one or more databases for molecules, reactions or data. Save and retrieve search results for use in future sessions.
ViewList Menu:	View search results. View tables of data. Create and combine lists of molecules or reactions.
Plot Menu:	Plot search results on monitor screen or to files for printing.
Forms Menu:	Create customised forms and tables to display reaction and molecule data.
Help Menu:	Displays the help pages and subtopics for the current menu.
Set Menu:	Change settings to customise features.

## Documentation

There are currently three manuals available for the program:

- REACCS Tutorial  
A step by step introduction that gives new users the basic information needed to start using REACCS.
- REACCS Searching Examples  
A set of searching examples describing search strategies for a variety of chemical questions. Concentrates on effective methods for formulating queries and conducting searches.
- REACCS Quick Reference  
A pocket size summary of commonly used commands, switches, and settings with a short explanation of each. Provides a quick reminder of the commands and their functions.

These manuals are sent out to everyone who registers for REACCS. They are also available on request through the online document ordering service ORDERDOC.

## Online help

Online information and help about REACCS is available on DLVB. Simply type R\$HELP at the \$ prompt to enter a VMS style help facility. If you are unfamiliar with this type of help, type README for more detailed instructions.

## Summary of databases available from the CDS

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

- CSSR** The Cambridge Structural Databank. Crystal structure data for about 115,000 organic and organometallic compounds. Currently runs under our in-house software, but we will make Cambridge's QUEST retrieval software available from 1 April 1994.
- ICSD** Inorganic Crystal Structure Data File. About 35,800 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 over 182,000 crystal structures.
- FNMR** A databank of about 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 11,000 entries.
- REACCS** A chemical reaction information management system allowing search, retrieval and display of molecules, reactions and their associated data. Currently contains about 212,000 searchable reactions and supplier information on over 129,000 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 66,000  $^{13}\text{C}$  NMR; 1,000  $^{15}\text{N}$  NMR; 850  $^{17}\text{O}$  NMR; 2,200  $^{31}\text{P}$  NMR; 1,800  $^{19}\text{F}$  NMR and 28,500 infra-red spectra. It does not, at present, contain any mass spectra.

## Terminal Emulation for SpecInfo and REACCS

This article contains helpful advice on setting up a suitable computer interface for SpecInfo and REACCS. Both packages are graphics based, so a terminal or emulator capable of supporting graphics is required. A package which only emulates VT100 or VT220 (such as kermit) is not going to work.

A large number of emulators are available and some work better than others. We have tested a variety of them and recommend the following:

VersaTerm Pro version 5.0.1 for the Apple Macintosh

Emu-Tek for the IBM-PC.

Chemtalk and Teemtalk for the IBM-PC will also work.

It is also possible to access both programs via X-terminals or X-terminal emulators. Emulators which we have found to work are:

MacX for the Apple Macintosh.

Vista eXceed for the IBM-PC.

Both Emu-Tek and Vista eXceed are available through the University CHEST agreement. Your University computer service should be able to provide details.

Some online help for terminal setups is available on DLVB. Use the R\$HELP and S\$HELP commands to access the help information for REACCS and SpecInfo, respectively. Typing R\$HELP or S\$HELP at the \$ prompt enters a VMS style help facility. If you are unfamiliar with this type of help, type README for more detailed instructions.

If you have any problems or need more details about terminals don't hesitate to contact us at the address given on the back page.

## CDS Training Courses

We plan to hold a regular series of training courses for the databases we run. These will be one day events held at a variety of regional centres as well as at Daresbury. The first three courses will be held at Daresbury in March. Each course will take up to 12 participants and will provide a hands-on introduction to one or more of the CDS packages. There will be no course fees, but we will not be able to meet travel and subsistence expenses.

March 15	General course
March 16	REACCS course
March 17	SpecInfo course

Future dates and locations have yet to be decided, but will be tailored to suit the needs of the user community. If you are interested in attending one of the above courses or would like a course to be held near you, please detach and return the form at the end of this newsletter.

## The Cambridge Structural Database on CDS

The Cambridge Structural Database (CSD) is currently available via the CDS component CSSR, using Daresbury's locally developed retrieval code. CSD is produced by the Cambridge Crystallographic Data Centre (CCDC) and is the world's definitive source of crystallographic data for organic and organometallic molecules.

In addition to compiling the data, CCDC have developed their own software system for search, retrieval, display and analysis of CSD information. The CSD software system is used world-wide by both academic and industrial research groups and is supported and developed on an ongoing basis by CCDC staff in Cambridge.

The CSD software system is accessed via a menu driven graphical interface. The various components support all existing bibliographic and 2-D substructure search facilities currently provided by the CSSR system, together with important new features, such as 3-D Geometrical Searching (QUEST3D), interactive Graphical Display (PLUTO) and Data Analysis (VISTA)

The CSD software system is currently available in-house at Daresbury. From the 1st of April we will make it available to all off-site users who currently have access to CSSR. The CCDC software will eventually supersede CSSR, but both systems will be available for a transitional period. It is planned to retain CSSR until

31 March 1995.

The CCDC will provide us with an Introductory Guide to the CSD software system and other training material, which we will distribute to users free of charge. A complete four-volume documentation set for the CSD system is available directly from the CCDC. We will be organising special training courses for the CSD system during 1994 and 1995.

For those attending the 1994 British Crystallographic Association Spring Meeting in Newcastle, there will be a demonstration of the CSD system given by CCDC staff, and a hands-on session. The BCA Meeting runs from Monday 28 to Thursday 31 March, with the CCDC demonstration on the Tuesday morning and the hands-on session, for restricted numbers, on the Wednesday morning. Registration before 12 February gives a discount in fees, whereas a large late fee applies to registrations after 5 March. For further details contact the local organisation:

Prof. W Clegg	Telephone	091 222 6649
Department of Chemistry	Fax	091 261 1182
University of Newcastle	E-mail	W.Clegg@newcastle.ac.uk
Newcastle upon Tyne		
NE1 7RU		

---

## CDS User Forum

All users of the Chemical Database Service are invited to attend the CDS User Forum, to be held at Daresbury Laboratory on Tuesday 17th May 1994.

The meeting will provide an overview of the current service as well as giving participants a chance to voice their concerns and needs.

A preliminary program for the meeting is:

- Talk: An overview of the service and the databases available
- Discussions: User needs  
Formation of a user group
- Demonstrations of databases

All users are invited, though those interested in forming a user group are particularly encouraged to attend. We will be able to pay travelling expenses for those attending the meeting, though attendance will be restricted to one member per research group.

Anyone interested in coming should contact Dave Fletcher at the address given on the back page.

# 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](mailto: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.

## Database queries:

Chemical Database Service (CDS)  
Room C18, Daresbury Laboratory  
Daresbury  
Warrington  
WA4 4AD  
(0925) 603162  
Email [cgsbb@daresbury.ac.uk](mailto:cdsbb@daresbury.ac.uk)

## For REACCS specific problems:

Dr. D. Parkin  
Email  
[D.Parkin@daresbury.ac.uk](mailto:D.Parkin@daresbury.ac.uk)

## For SpecInfo specific problems:

Dr. D.A. Fletcher  
Email  
[D.A.Fletcher@daresbury.ac.uk](mailto:D.A.Fletcher@daresbury.ac.uk)

## For general problems contact the Service Manager:

Dr. R. F. McMeeking  
Email  
[R.F.McMeeking@daresbury.ac.uk](mailto:R.F.McMeeking@daresbury.ac.uk)

## Documentation:

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

## Access:

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

DTE address:  
000001003000

NRS name:  
UK.AC.DARESBUY.DLVB

Internet number:  
148.79.160.7

Internet name:  
[dlvb.dl.ac.uk](http://dlvb.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](mailto:D.A.Fletcher@daresbury.ac.uk)