Chemical Database Service

NEWSLETTER

Issue 8 November 1997

As a result of the recent successful trial, the chiral separation database, **Chirbase**, has been added to the CDS portfolio. This decision will be reviewed after a year. More information about Chirbase, and the database access files needed to access the data via ISIS, are available from the CDS website.

Two **solid-phase synthesis** databases are currently on trial, and we have plans for future trials of a number of databases and software packages. See the articles on pages 2 and 5 in this newsletter for more details.

The ChemInform RX database has been updated under **ISIS**. Users will need to update their database access files in order to use this new data. See page 6 for details of how to do this.

We have just learnt that from April 1998 MDL will no longer enhance or release its **ISIS/Client** software in the 16 bit Windows or 68k Macintosh operating systems. In the long run this will mean users will need to migrate to Windows 95 or NT (Windows 3.1 with Win32s will not be supported) or PowerPC Macintosh, though we suspect that current versions of ISIS/Client will continue to work for some time to come.

A CHEST deal has been agreed for the Macintosh X-Windows emulator package **eXodus**. Single copies are available for only £30. See page 8 for details.

CDS will be upgrading its server at the end of this year. Some details of the new machine and timescales for its introduction are given on page 10.

An basic **SpecInfo** course covering structure drawing, spectrum prediction and data output, and lasting about 1/2 a day is now available. If you are interested in attending or would like us to run this course at your site, please contact Dave Fletcher at the address given on the back of this newsletter.

Contents

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

Solid-phase synthesis database trials

Two new databases in the area of solid-phase synthesis have been obtained on a trial basis until the end of January 1998. Solid-phase synthesis has major relevance for the synthesis of combinatorial libraries used in drug discovery research. Solid-phase synthesis databases provide ready access to up-to-date information in this important and rapidly-developing area of chemistry.

The two databases are described below and details of how to access them are given at the end of this article.

Solid-Phase Organic REactions (SPORE)

SPORE is a new database from FIZ CHEMIE and made available through MDL. An ISIS application is included with the database to facilitate accessing SPORE's unique information, including multistep schemes, polymer supports, solid supports, and protecting groups. The database contains more than 4,100 reactions from 250 journals and is updated quarterly.

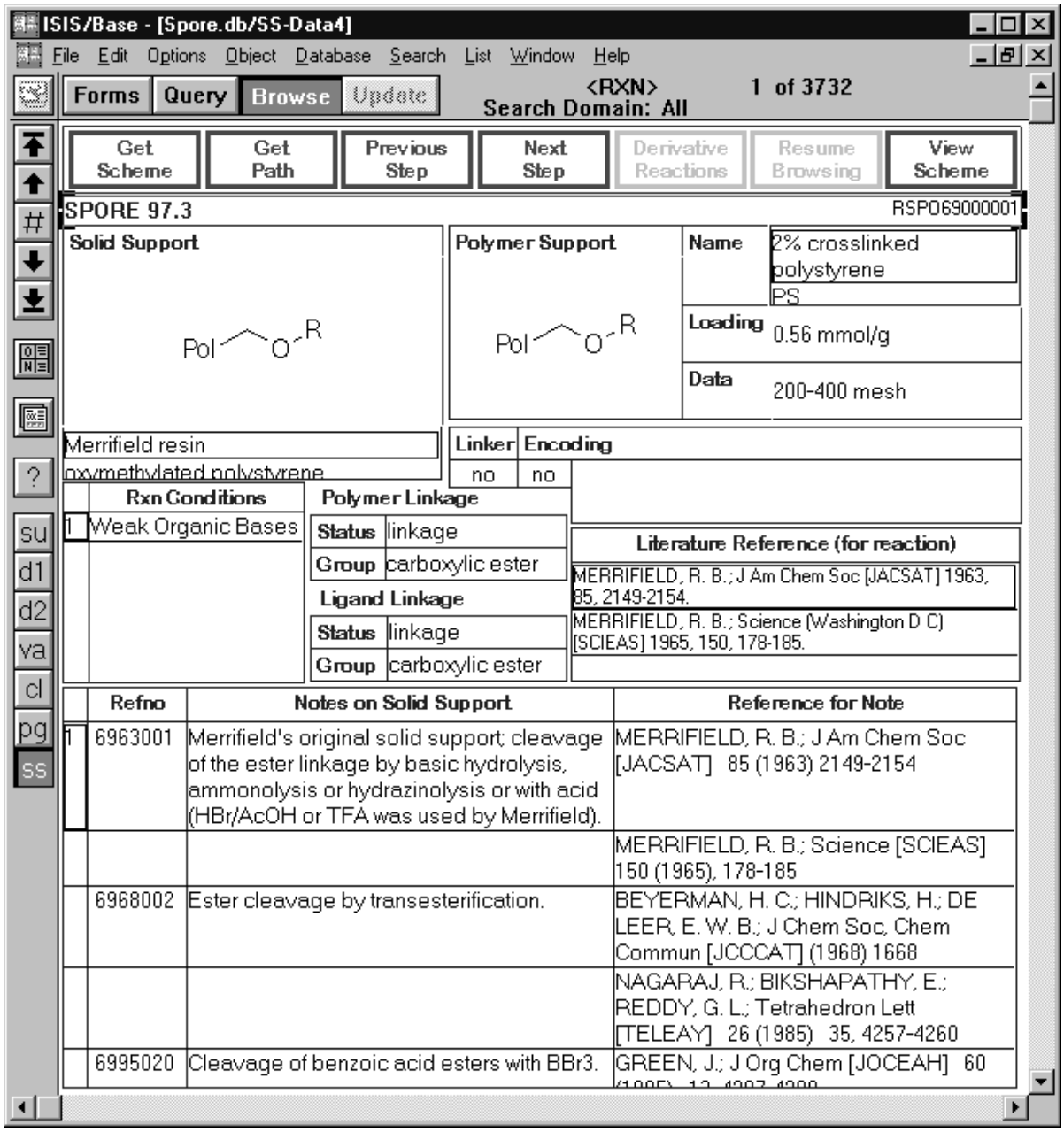
The familiar browser form contains boxes for structures, reactions and data, but the data box has been modified to contain fields useful for searching this particular database. Clicking in one of the data boxes brings up the field vocabulary box where you are presented with a list of possible field values.

SPORE uses the unique clustering feature of ISIS, where the hits can be clustered according to one of a number of factors such as reaction conditions. A number of different forms are also available for viewing the hits. As well as the usual summary, data, variations and cluster forms, two new forms are available, the protecting groups data (pg) form and the solid support data (ss) form. These forms can display schemes and paths like the d1 and d2 data forms in the RXL Browser.

SPORE is not currently available for the SGI.

Solid Phase Synthesis (SPS)

The Solid-Phase Synthesis database from Synopsys contains a selection of relevant material from primary literature, reviews and patents. Selection of the data is carried out by experts at Oxford Diversity Ltd. and abstraction of the literature, compilation of the database and quality control is carried out by Synopsys. The database is sourced from over 1,000 publications, and contains nearly 5,500 reactions. It is updated quarterly.

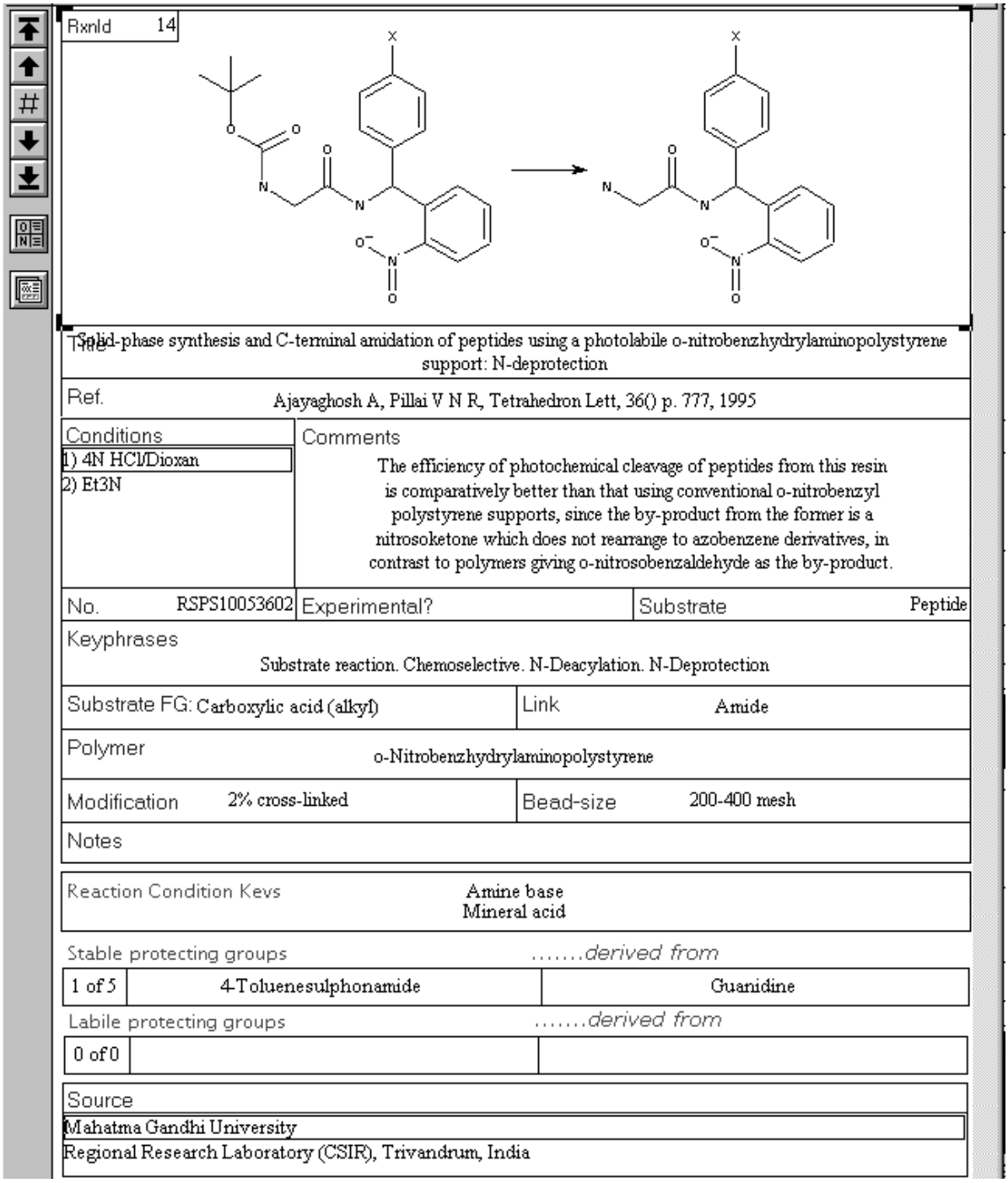


SPORE browser screen shot showing a retrieved entry from the database

The focus is on chemical transformations, selected to illustrate the variety of reaction types which have been demonstrated on the solid phase. Emphasis is placed on 'small-molecule' transformations but 'state of the art' methods and new developments in the areas of peptide, oligosaccaride and oligonucleotide synthesis are also covered.

The SPS database contains 3 forms for searching and viewing the data.

Catsol form - used for displaying the catalysts and solvents.
Viewmol form - used for viewing the reactant and product molecules.
Default form - main form which contains most of the fields.



An entry from the SPS database shown in the default form

How to access the trial databases

In order to access these trial databases, you will need to download new database access files from us. The simplest way of doing this is via the CDS website (http://www.dl.ac.uk/CDS/cds.html). Alternately the files can be downloaded via ftp from cds1.

Database access file	File type	File
SPS for PC	Binary	/ftp/isis/ibm/sps.db
SPORE for PC	Binary	/ftp/isis/ibm/spore.exe
SPS for Mac	ASCII	/ftp/isis/mac/sps.hqx
SPORE for Mac	ASCII	/ftp/isis/mac/spore.sea.hqx
SPS for SGI	Binary	/ftp/isis/sgi/sps.db
SPORE for SGI	not available	

Informed user opinion will be critical in determining whether either of these databases become part of the regular CDS service.

Future trials

Database trials are an important part of the process of obtaining new data sets for CDS. In general we would expect to trial any new system or database and survey user opinion before adding it to the CDS portfolio. The recent trial of the chiral separation database, Chirbase, was an example of this procedure.

Currently there are two solid phase synthesis databases, SPORE and SPS, on trial, accessible via ISIS (see page 2 for more details).

In the future we are expecting to run trials of the following ISIS databases:

Metabolite from MDL

Metabolite is a complete metabolism information system including key information about metabolic transformations of xenobiotic compounds from the literature. Information in the database was abstracted from; Biotransformations of Drugs (1977-1983), Pharmacokinetics (1986-1990) and Original metabolism literature (1990 onwards). MDL updates Metabolite semi-annually. The database indexes path and scheme information for its metabolic transformations and includes such additional data as; species, route of administration and excretion, analytical methodology, parent compound class and physiological activity.

Biocatalysis from Synopsys

BioCatalysis has been developed in association with Professors Bryan Jones and Herbert Holland to recognise the increasing importance of biomolecules as catalysts in organic synthesis. It focuses on the synthetic use of enzymes and micro-organisms which offer advantages of excellent chemo-, regio- and enantioselectivity, coupled with important environmental benefits. Version 97.1 comprises 19,500 reactions, abstracted from almost 8,000 citations (including patents).

Bioster from Synopsys

The BIOSTER database, from Synopsys, is a critical compilation of over 2000 bioisosteric transformations (including drugs, agrochemicals and enzyme inhibitors), selected from several thousand references published over the last 35 years. The database provides keywords indicating the mode of action and species the bioanalogs were tested against. Each transformation provides cross-referenced information to the papers in which it was reported.

Metabolism from Synopsys

The database comprises metabolic transformations of chemicals, including pharmaceuticals, agrochemicals, food additives and industrial chemicals, in mammals, birds and fish. Version 1.0 of the database comprises 9,300 transformations of 1,916 parent compounds, abstracted from over 1,900 primary literature sources between 1987 and 1994. It is derived from the series 'Biotransformations', published by the Royal Society of Chemistry in the UK.

We will also be running trials of two new database access systems, once migration to the new CDS server is complete (early 1998). These are:

Chemscape from MDL

This provides web access to the organic reaction databases, current available through ISIS. It has similar functionality to searching remote databases via ISIS/Base, but is run through a web browser equipped with the Chime Pro plug-in.

SpecInfo/ISIS link

This allows access to the SpecInfo package from ISIS/Base. We expect to be able to run a two month trial of this system early next year.

ISIS - are you searching all of the data?

The ChemInform RX database currently contains over 430,000 reactions and is growing at the rate of 70,000 reactions per year. 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 will not be searching the most recent reaction data.

Statistics for September 1997 show that 5% of all searches in the ChemInform database were not accessing the data added in 1995 and nearly 20% of searches were missing the data added in 1996.

This year's update to the ChemInform RX database has just been installed (November 97), 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 RXL Browser file from the RXL Browser folder.
  - PC Select the file rxl132 from the rxlbro directory.
  - SGI Select the file all.db.
- Once the database is open select Retrieve All from the Search menu.
- ISIS should retrieve a hitlist of some 710,900 reactions. If you get fewer reactions you are not accessing all of the data.

How to update your ISIS database access file (for PC and Mac only)

- 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 RXL Browser file from the RXL Browser folder.
  - PC Select the file rxl132 from the rxlbro directory.
- Check the large box, it should contain the following hview list (the order is not important) :

  reflib+orgsyn+chc+jsm+cirx92+cirx93+cirx94+cirx95+cirx96+cirx97+cirx98

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

Note that SGI users cannot update their ISIS database access files by this method, and will need to download a new file (see next section for details).

How to download a new ISIS database access file

You can download new ISIS database access files from the CDS website at http://www.dl.ac.uk/CDS/cds.html. Choose the Information for Users link and then the Download link. 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 (RXL Browser link for Mac and PC, Main reaction databases link for SGI).

Alternately you can download the relevant files via ftp from cds1.

Machine	File type	File
PC	Binary	/ftp/isis/ibm/rxl132.db
Mac	ASCII	/ftp/isis/mac/RXL_Browser.hqx
SGI	Binary	/ftp/isis/sgi/all.db

The Mac file needs to be unbinhexed and placed in the RXL Browser folder in the ISIS folder. The PC file needs to be placed in the rxlbro directory in the ISIS directory.

X-Windows emulators

An X-Windows system is needed to access SpecInfo and is useful for Quest and the other Cambridge Structural Database System (CSDS) components. All systems, at present, can be accessed using X-Windows, whereas terminal emulators (e.g. Tektronix 4105 emulators like VersaTermPro or EmuTek) are becoming less useful. They are no longer supported by SpecInfo and some display utilities, and in the future will not be supported by Quest and the other CSDS components.

Web based access to all of the CDS data is some way off, therefore X-Windows is currently the best way to access all of the CDS data (excluding the organic reaction data, which is best accessed via its own ISIS software). For most people this will mean running an X-Windows emulator on their PC or Mac.

Current X-Windows emulators available at good rates for UK academics.

MI/X

For 68k Macs, PowerMacs, Windows 95 and NT PCs.
It is available free and can be downloaded over the web from http://tnt.microimages.com/freestuf/mix.htm. Look in the FAQ file for European mirror sites for faster downloading. Unfortunately this emulator does not work with SpecInfo at present.

eXodus

For 68k Macs and PowerMacs.
A recent CHEST deal means that it is now available for £30 per copy (electronic download, no manual or support) from E92plus. Contact

Darren Parker, Education Account Manager at E92plus
Tel: 0181 399 3111
Fax: 0181 399 5111
Email parkerd@e92plus.co.uk

MacX

For 68k Macs and PowerMacs.
This is also available through CHEST as part of the ClarisPlus Master Volume Licensing Agreement for £23-84 per copy. See http://www.chest.ac.uk/software/claris/contents.html for more details.

eXceed

For Windows 3.x, 95 and NT PCs.
Site licenses available from CHEST, £825 - £1700 per site depending on number of operating systems supported. See http://www.chest.ac.uk/software/ameridata/contents.html for more details. Many sites are already subscribed to this deal.

X-Win32

For Windows 3.x, 95 and NT Pcs.
A demonstration version, which only works for 2 hours at a time can be obtained from http://www.starnet.com. 16 bit Windows 3.1 and MS-DOS versions are also available. Full versions cost $200, though academic discounts are available.

Our current recommendations are eXodus for Macs and eXceed for Pcs.

Report on this year's User Forum

This year's meeting was held on July 9th 1997 at Daresbury Laboratory. Much of the information and discussion is reported in other articles in this Newsletter. A few additional points were raised/mentioned:

- There was strong support from the meeting for a continuation of the CSSR service.
- The upgraded CDS web site now includes downloading of documentation and client software, online manuals and newsletters, a database of common problems and an online registration procedure.
- CDS may produce a simple web interface for the crystallographic data.

Future meetings

Low attendance seems to be the norm for these meetings. We intend to run a survey to find out the preferred options for timing, location and frequency of future meetings. Other options which might be considered include:

Stopping regular user meetings
Having meetings alternate years
Having meetings at different locations each year
A combined user meeting with MIDAS/CrossFire

Upgrade of CDS hardware

The current CDS hardware is reaching the stage where maintenance costs and reliability start to become an issue. Cds1 is nearly four years old, and its annual maintenance cost is a significant fraction of the cost of a similarly powerful machine. The recent successful refunding bid for the service included provision for a hardware upgrade. The current plan is to replace cds1, cds2 and cdssg1 with a Silicon Graphics Origin 200 server.

Moving to a Silicon Graphics system will allow us to trial two new software systems (MDL's Chemscape - web access to the ISIS databases and the SpecInfo / ISIS link - allowing access to the SpecInfo package from ISIS / Base) - see the article on forthcoming trials for more details.

There will be no changes to the operation of the current CDS database packages. The main effect of the hardware change on users will be the need to reregister on the new machine. This will probably be handled via a web based form similar to the MIDAS registration system for Beilstein. This has the added advantage of cleaning up old ids, so there will be no annual "dead ids" exercise this year.

The tender exercise to purchase the new server is now under way and we expect to have the new machine installed before the end of the year. We would hope to start moving users onto the new machine early in the new year and anticipate a 3 month overlap period when the cds1 service and the service on the new machine are both available. At the end of this period, access to cds1 will be stopped and all users, who have not already done so, will have to move to the new machine.

Tentative timescale for CDS hardware upgrade

Install new server	December 97
User access to new server	Jan-Feb 98
End parallel service	April-May 98

Specifications of new CDS server

- Silicon Graphics Origin 200 Server running IRIX 6.4.1
- 2x R10,000 180 MHz CPUs
- 256 MB RAM
- 3 x 9 GB disks
- CD-ROM drive, other drives from current CDS hardware

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 730,000 searchable reactions and supplier information on 250,000 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}C NMR; 999 ^{15}N NMR; 856 ^{17}O NMR; 2,183 ^{31}P NMR; 1,825 ^{19}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 175,000 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 6,500 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}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://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 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