the Chemical Database Service

NEWSLETTER

Spring 2010

Contents

Welcome

Welcome to the Spring 2010 Newsletter of the Chemical Database Service (CDS)

See inside for details of the current state of the Service and what has changed over the past 6 months.

Latest News

Login/Password/Registration

A number of changes have been made.

More details on page 2.

Updates to Databases

See the latest news and updates

More details on pages 4 and 5

Accelrys Databases

CDS now make the Accelrys databases Protecting Groups, BioCatalysis and Solid Phase Synthesis available again

More details on pages 6 and 7.

Symyx Trial

An evaluation trial is available for all of the Symyx databases including Synthesis, Sourcing and Bioactivity databases.

Try out the Available Chemicals Directory and let us know what you think.

More details on pages 8 and 9.

Science & Technology Facilities Council

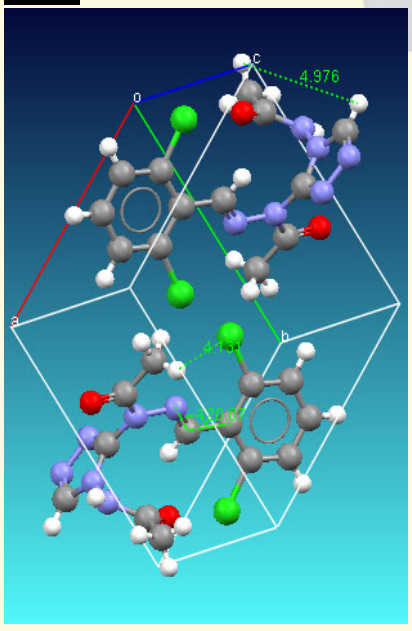
EPSRC

Links used in Newsletters can be found at:- http://cds.dl.ac.uk/letterlinks

Crystallography Databases

ICSD

CrystalWeb



Thermophysical Database

DETHERM ... *on the WEB*
Thermophysical Properties of Pure Substances & Mixtures

ACD/I-Lab

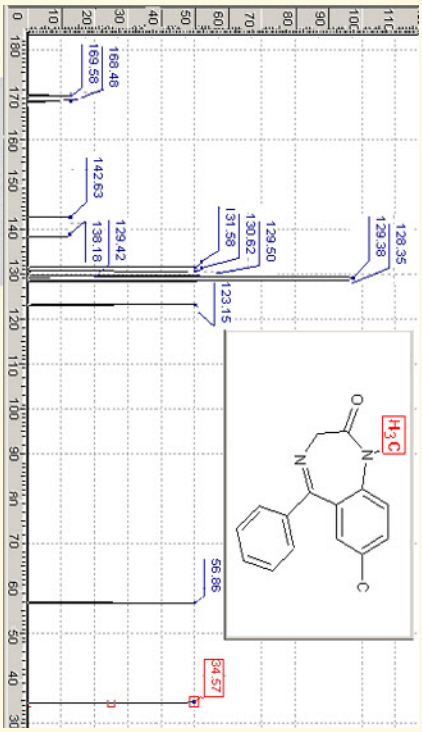
1 pKa
2 Search pKa DB
3 LogP
4 Search LogP DB
5 LogD
6 Aqueous Solubility
7 Search Solubility DB
8 Boiling Point/Vapor Pressure
9 Adsorption Coeff./BCF

Boiling Point:
BP (deg C at 760.00 mmHg) = 431.15 ± 18.00

Vapor Pressure:
$\log_{10}(p) = -8.516 \pm 2.33$ (p in mmHg)
p (at 25.00 deg C) = 0.0000000030 mmHg

Enthalpy of Vaporization:
$\text{delta}_{vap}H(T_{boiling\ point}) = 79.29 \pm 6.0$ kJ/mol

Spectroscopy Databases



News

Logged in Status

Following suggestions by our users,a Status line is now displayed to show if you have logged in or not.

[Status : not logged in]

In the future we will make this more clear - possibly by removing the username and password boxes completely once logged in.

I forgot my password

If you have lost your password, you can now generate a replacement one online. All you need to know is your ID. Simply click on '**I forgot my password**' just under the sign on boxes.
Its also available via the "Registration" pull-down menu on all CDS web pages, under "Lost Password or unsubscribe".

On submission, an automatic email will be sent to your registered email address, containing a web link. This link will generate a temporary password for you when opened in a web browser which you can then change (to something more memorable) via the "Registration" pull-down menu on the home page, under "Change your Password" [or directly at https://cds.dl.ac.uk/pass.html]

Institutional logins

There was a mandatory upgrade of Shibboleth 1.3 to 2.0. Users should be able to log in using their Institutional ID's provided that the Institutional ID has been mapped with the CDS ID.

Twitter

News items are now available on twitter. **cds_daresbury** has been set up for you to follow and obtain the latest information or you could just select the link from the CDS home page. [You do not have to register with Twitter]
There is also a CDS Twitter RSS feed available

Helpdesk Email Address:

We have changed the email address for the CDS Helpdesk from cdsbb@dl.ac.uk to

cdsbb@stfc.ac.uk.

This is to conform to a global policy for all STFC sites. All "dl.ac.uk" email addresses will cease to work at some point in the future. The address cdsbb@dl.ac.uk will continue to work for a period, but we urge you to update your records to use the new address. We are changing the information on the CDS webpages to reflect these changes. There are, however, no plans to change addresses for the CDS website and for direct logins to the main CDS server (cds.dl.ac.uk).

News

RIN/JISC Survey

A study was commissioned by the Research Information Network (RIN) and the Joint Information Systems Committee (JISC) to explore the long-term use and benefits of Research Data Centres.

The study featured the activities of a small number of well-established UK Data Centres such as CDS and aimed to document the benefits and impacts that effective sharing of research data can deliver to the scientific community and other users.

At the beginning of January 2010, a survey was carried out and a quick summary of the results are shown below:

- Users access the data from CDS primarily for their own use.
- Some users have been accessing data for over 20 years and, of the 200 users that replied, they have obtained data from CDS, on average, around 450 times. 75% of those that replied said they accessed data weekly or monthly while 4% accessed it daily.
- 30% of users said that the extent to which they access data from CDS has increased over time with 54% of users remaining the same.
- The type and form of the data accessed varies but Experimental data in the form of processed data is the largest.
- Most users use the data obtained from CDS for combining with other data and use it in research papers for publication.
- 24% of those that replied had never cited the Chemical Database Service, the rest were split between sometimes (42%) and always (34%).
- The importance of the data available from CDS for your research was stated as **essential** by 34% of those that replied while others said it was quite (28%) or very important (30%).
- Users were of the opinion that the same data would be available elsewhere, but would be more difficult and more expensive to access. Over 12 % thought that same data would **not** be available elsewhere, and no usable substitutes would be available.
- On average, **over 50%** said that accessing data via CDS had benefited their research to a **large extent** with 34% saying it had benefited to a small extent (such as improved the efficiency of research)
- 79% of users are satisfied or very satisfied with the Service as a whole (with 19% remaining neutral in their opinions).

News and Updates

Crystallography

Cambridge Structural Database System

500,000th Crystal Structure Milestone Reached
For more details see: http://www.ccdc.cam.ac.uk/500000.php

CSD Update

Version CSD V5.31 (2010 Release)

The latest update (added March 2010) includes 7,380 new entries. The total number of entries is now 503,348.

This CSD Update also flags a number of retractions for some entries in the existing database that correspond to the recent article by IUCr in *Acta Crystallographica Section E: Structure Reports Online (2010), E66, e1-e2*, which cited the publication of 70 structures based on falsified data.
See http://www.ccdc.cam.ac.uk/retraction.php for more details.

ConQuest: Searches can now be restricted to just the best representative of each unique polymorph, thus eliminating hits from duplicate crystal structures. Four best representative subsets are available based on R-factor, the presence of hydrogens, room temperature and lowest temperature determinations.

Mogul now contains a library of ring conformations, allowing you to assess how the geometry of a ring in a query structure compares with that of similar rings in the CSD. A single figure measure of geometrical similarity is calculated and a histogram containing matching CSD rings displayed. This makes it easy to identify if the query ring has an unusual geometry.

Mercury 2.3: The **Materials module** has been improved and has a new wizard which provides access to the range of different tools for performing interaction analyses and packing similarity calculations.
You can now apply crystallographic or non-crystallographic symmetry operators to molecules using the new **Transform Molecules** tool.

Other additional features include:

- Improved Molecular overlays - change space group settings at the click of a button
- Change space group settings easily,
- Improved slicing tools giving easier visualisation of specific structural planes or crystal surfaces.
- Crystal structures can be coloured by symmetry operation
- A new display style has been introduced named "ORTEP" which makes the generation of publication-quality images with anisotropic displacement parameters a lot easier.

ICSD Update

The latest ICSD update (2010/1) contains 8,159 more entries as compared to Version 2009/2. The total number of entries is now **132,526.**

It can be accessed via the ICSD-WWW interface or via CrystalWeb.

Links used in this Newsletter and others can be found at:- http://cds.dl.ac.uk/letterlinks

News and Updates

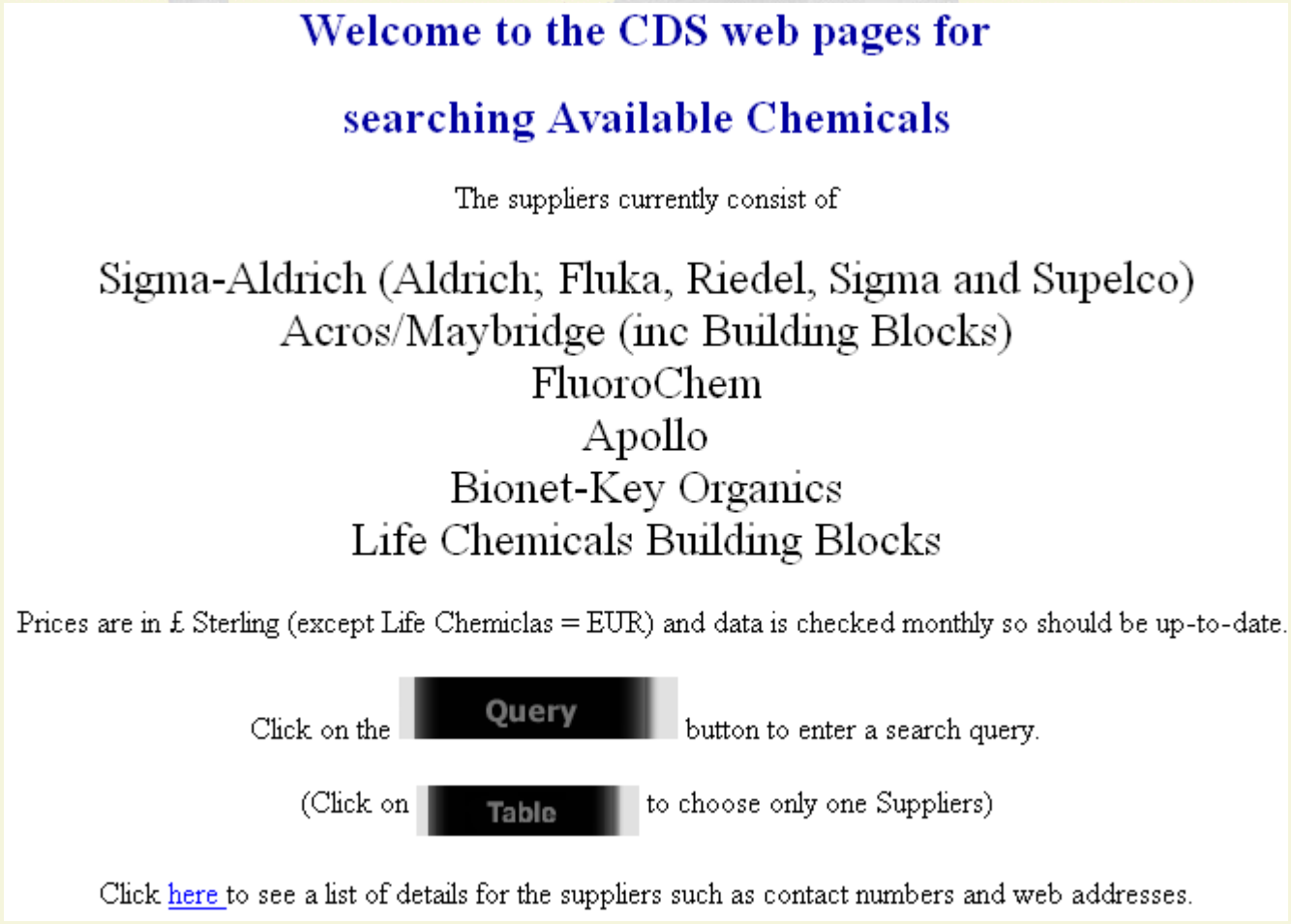
CDS Available Chemicals

Updates

The Suppliers database has been updated. Up-to-date prices have been maintained for the Aldrich set of databases plus Apollo and FluoroChem.

New Data

New data containing prices has been added for Key Organics [Core Scaffold (245 Compounds - prices for 500mg, 1g, 5g and 10g) and Screening Compounds (42,683 Compounds - prices for 1mg, 5mg and 10mg)]
Building Blocks from Maybridge (6,605 compounds) and Life Chemicals (6,595 Compounds - Prices for 500mg, 1g, 5g and 10g) have also been added.
All prices are shown in £ sterling apart from Life Chemicals which is in Euros.
This brings the total amount of compounds to 267,675.



New links have also been added to give contact details for the Suppliers

Screening Compounds

The Screening Compounds database has been updated with all the latest information from Suppliers and a new one added (Zelinsky Institute). This brings the total number of entries to 3,577,324.

Organic Chemistry

Accelrys Databases

The following databases are now available using the Accelrys Database Explorer client software for PC's (only)

* **Protecting Groups**
* **Solid Phase Synthesis**
* **BioCatalysis**

Accord Database Explorer (ADE) is a simple yet powerful **Windows** desktop application that enables you to work with the Accelrys chemistry databases. Using Accord Database Explorer you can:

- Search and view all three databases simultaneously
- Create queries with up to 10 search criteria.
- Sort and filter query hitsets according to their content
- Create reports to summarize hitset results

Accord Database Explorer features:

- Atom mapping - for true reaction searching
- Sophisticated query building - combine multiple search terms with Boolean operators
- Advanced hitset handling - save, combine and browse hitsets
- Compatible with leading structure editors - including ISIS/Draw, ChemDraw and CAS Draw

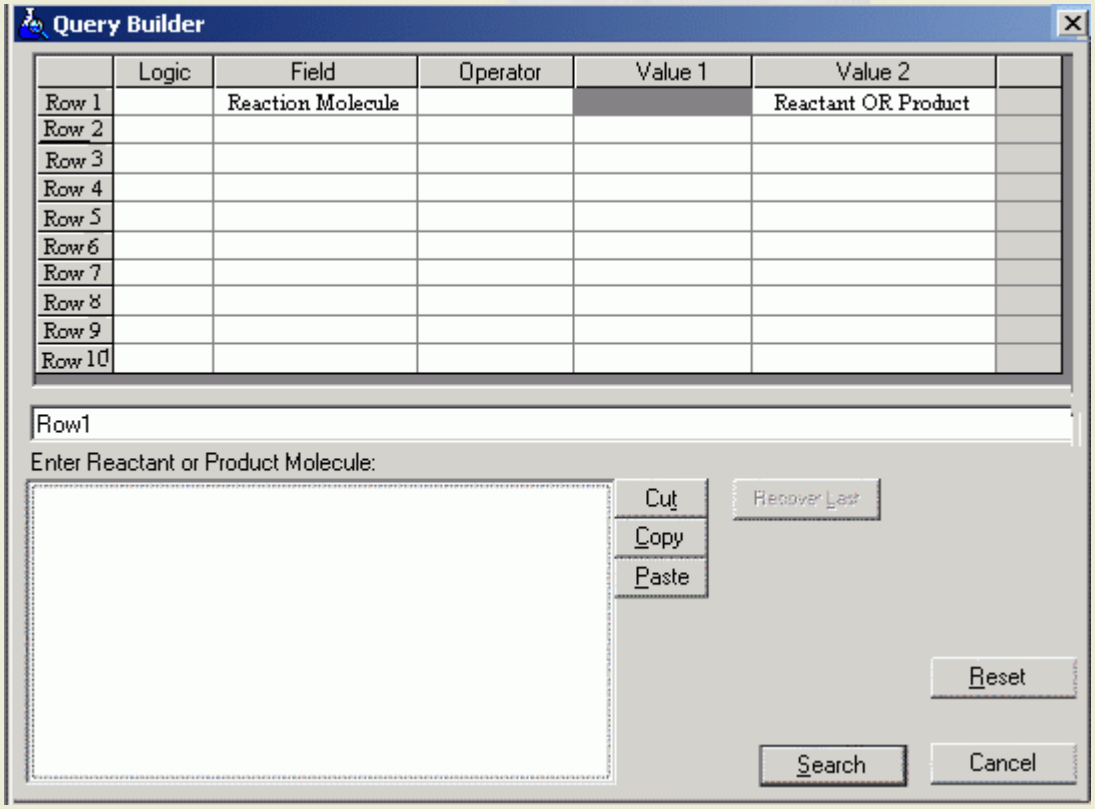
Searching the Databases

Use the **Hitset Manager** to Create, Import, View, Export, Remove and Sort Queries and Hits.

There are two main ways of searching:

Query Builder **Query By Form**

Query Builder Form



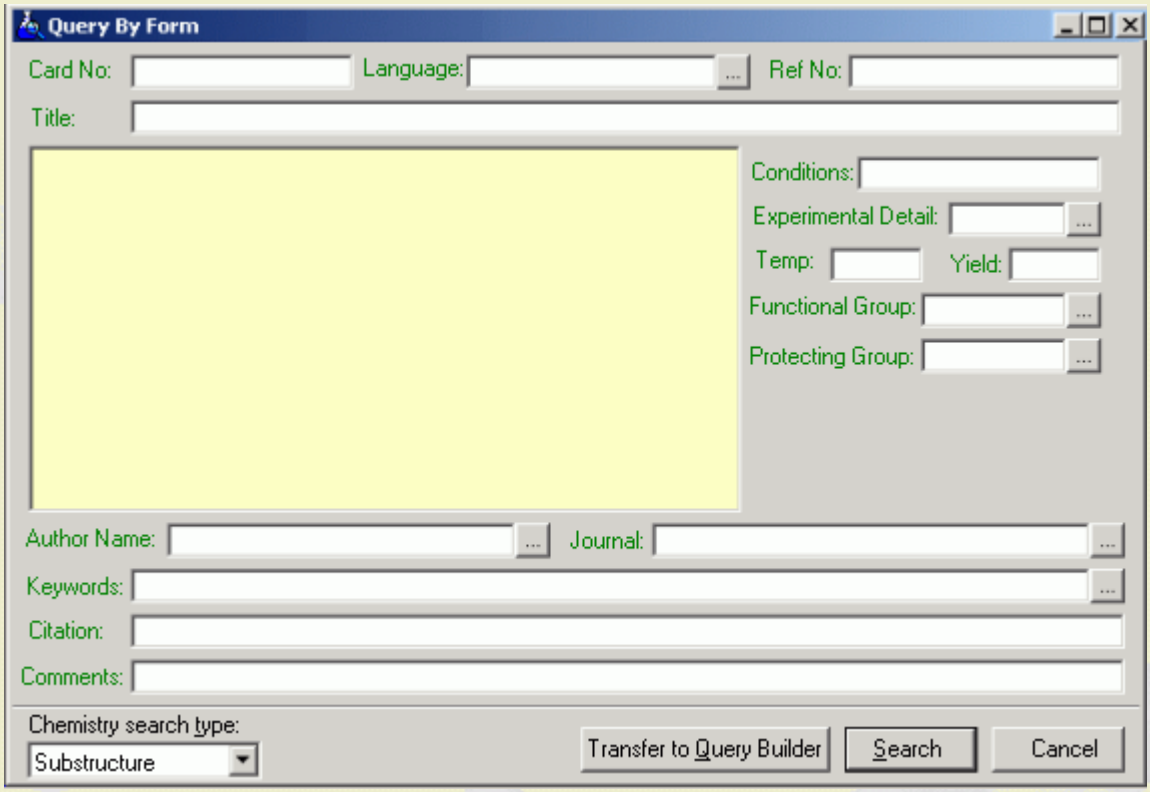
Create queries with up to 10 criteria (including structure searches) and combine them using logical operators (and, or, not)

Links used in this Newsletter and others can be found at:- http://cds.dl.ac.uk/letterlinks

Organic Chemistry

Accelrys Databases (cont.)

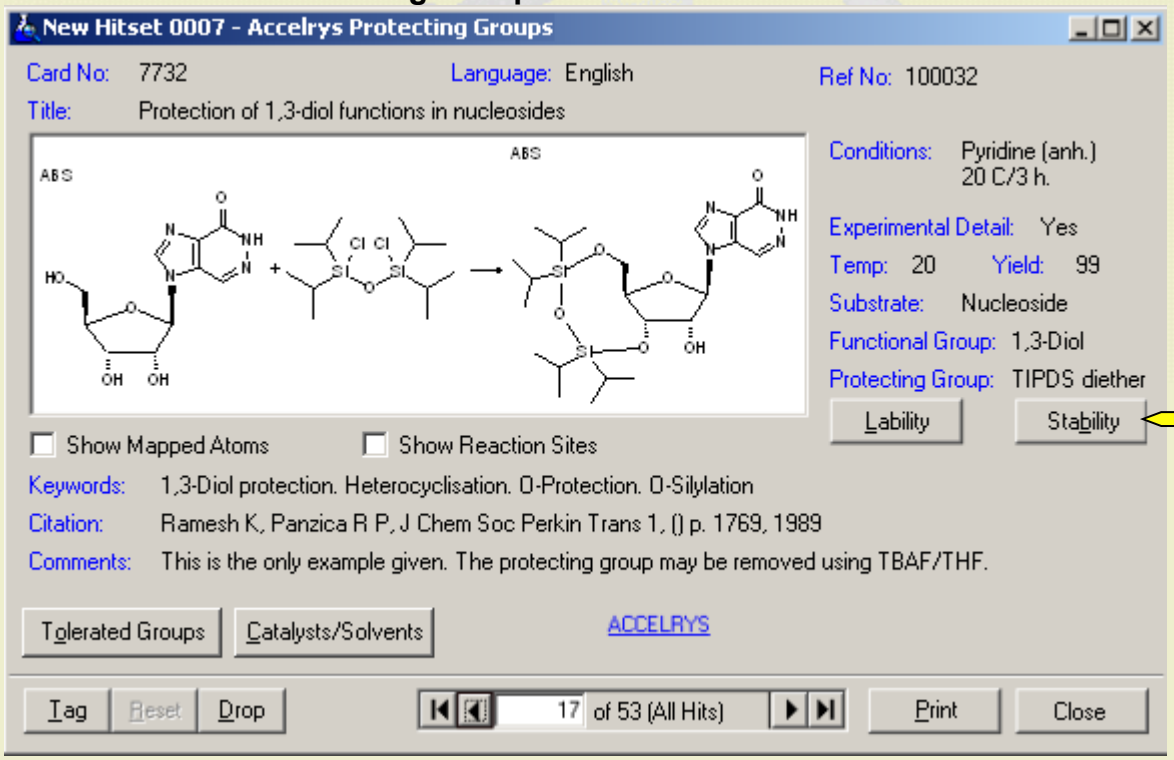
The Query by Form button displays a query form which is dependant on the database to be searched.
E.g. **Protecting Groups Query Form**.



Viewing the Results

The databases automatically display the results in the **Default Form** for that database as soon as a search is finished.

Results Form for Protecting Groups Database



Other results forms are available by clicking on the buttons

Links used in this Newsletter and others can be found at:- http://cds.dl.ac.uk/letterlinks

New Trial

Organic Chemistry

Symyx/Isentris Databases

CDS, in conjunction with JISC, are looking to re-establish the databases that we used to offer in the past using ISIS software, but will now be based on **Isentris** software.

CDS are running an evaluation trial for 3 months.

All databases from Symyx (see below) will be made available over this trial period. Access will be via an Isentris client and via DiscoveryGate.

This is not only your chance to evaluate these databases but also to send a message to EPSRC that these databases should be included in the next tender - so sign up and use them.

Databases

Bioactivity

* Comprehensive Medicinal Chemistry (CMC)
* MDL Drug Data Report (MDDR)
* Metabolite
* National Cancer Institute Databases - The NCI 127K, Plated Compounds, AIDS, and Cancer databases
* Toxicity - A compendium of toxicological data.
* RTECS (Registry of Toxic Effects of Chemical Substances)

Sourcing Databases

* Symyx Available Chemicals Directory (ACD)
* Symyx Screening Compounds Directory (SCD)

Synthesis Databases

* ChemInform Reaction Library (CIRX)
* Current Synthetic Methodology (CSM)
* Derwent Journal of Synthetic Methods (DJSM)
* ORGSYN Database
* Reference Library of Synthetic Methodology (RefLib)
* Solid-Phase Organic Reactions (SPORE)
* SPRESI (also available now via CDS)

More information can be found at:
http://www.symyx.com/products/pdfs/dg-discovery-package-ds.pdf

Links used in this Newsletter and others can be found at:- http://cds.dl.ac.uk/letterlinks

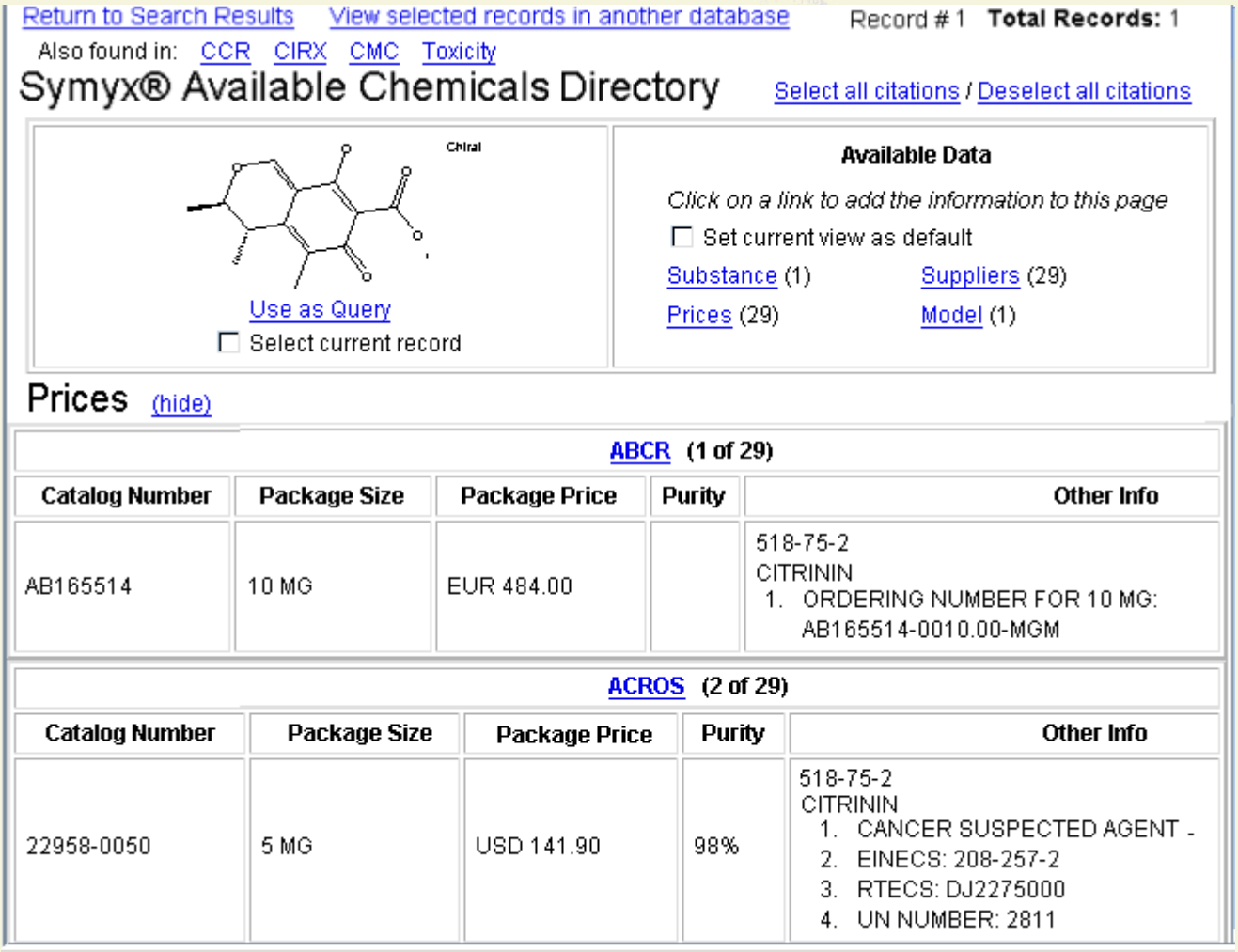
New Trial

Symyx/Isentris Databases (cont.)

Initially **users must contact CDS** and they will then be sent login credentials to access the databases via DiscoveryGate. It is hoped that sentris client software will be available later Ito access the data.

DiscoveryGate requires that Java 2 Runtime Environment (JRE) 1.5.0_12 or later is available on your PC and will install it if not found.

The Java based **Symyx Draw** is also made available on your desktop machine for entering structural queries.



A hit from Symyx Available Chemicals Directory

Symyx ACD contains pricing and supplier information for over 1,160,000 unique chemicals, including 3D models, from 840 suppliers worldwide.

Getting started information is available at:
http://cds.dl.ac.uk/cds/datasets/orgchem/symyx/discoverygate.html

Funding details of the databases has still to be worked out, but it is likely that the following 'packages' would be offered on an Institutional or UK wide model.

'**Discovery** Package' which includes all the Bioactivity, Sourcing and Synthesis Databases
'**Synthesis** Package' - ACD plus Synthesis Databases
'**Pharmacology** Package' - all the Bioactivity databases.

Links used in this Newsletter and others can be found at:- http://cds.dl.ac.uk/letterlinks

Research Highlight

Crystal structure prediction of carbon nitrides and phosphides

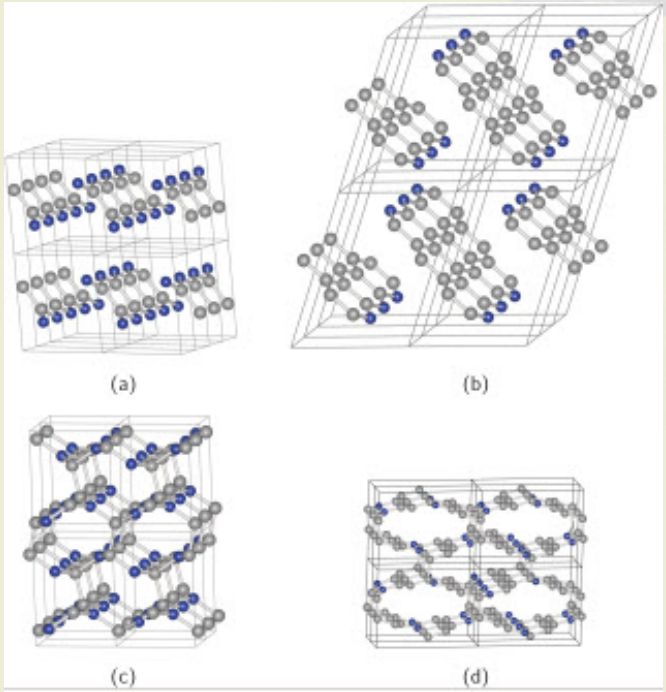
Judy Hart, Fred Claeyssens, Neil Allan, Paul May

School of Chemistry, University of Bristol

Developing the ability to predict crystal structures is an important problem in materials science, since this ability would allow for the rational design of materials with properties adapted for particular applications.

Carbon nitrides are an important group of materials and have been of much interest since the prediction that the β-C_3N_4 phase should have a hardness greater than diamond [1]. However, synthesis of this material has not yet been successful. The products of synthesis often have a high carbon:nitrogen ratio, suggesting that the incorporation of nitrogen into carbon is difficult. Therefore, we have investigated the structures and properties of carbon nitrides with a relatively high carbon content, such as CN and C_3N.

In doing this, we have developed an approach to crystal structure prediction that emphasises the local bonding enviroments in the crystal. Our approach involves starting from the crystal structures of known binary compounds.



Predicted low energy structures for (a) CN and (b) C_3N at low pressures, and (c) CN and (d) C_3N at high pressure.

The CDS is a valuable service for allowing the rapid collection of a range of plausible starting structures. We then conduct full geometry optimisation of these structures using density functional theory (DFT) calculations. Based on the results, it is often possible to rationalise their relative energies in terms of, for example, known bond enthalpies, whether atomic valencies are satisfied and the presence of strain. By exploring a range of compositions, conclusions can be drawn regarding the preferred local bonding environments for each element in the compound as a function of both composition and synthesis conditions such as pressure.

For carbon nitride, we have successfully predicted crystal structures across a range of compositions [2]. Under positive hydrostatic pressures, structures derived from the β-InS structure, in which the nitrogen is three-coordinated and the carbon is sp^3 hybridised will be preferred. At low pressures (or high temperatures), graphitic structures with some electron delocalisation will be preferred.

We have similarly undertaken studies to predict the crystal structures of carbon phosphide as a function of composition [3]. Graphitic structures with formal electron delocalisation are preferred when the phosphorus content is low, but when the phosphorus:carbon ratio is ≥ 0.2, β-InS-like structures with hypervalent, four-coordinate phosphorus and sp^2 hybridised carbon are preferred. When the phosphorus:carbon ratio is > 0.5, the lowest energy structures are pseudocubic-like and contain sp^3 hybridised carbon and three-coordinate phosphorus.

References

1. A. Y. Liu and M. L. Cohen, Science, 1989, **245**, 841–842.

2. J. N. Hart, F. Claeyssens, N. L. Allan and P. W. May, Phys. Rev. B, 2009, **80**, 174111.

3. F. Claeyssens, J. N. Hart, N. L. Allan and J. M. Oliva, Phys. Rev. B, 2009, **79**, 134115.

Databases Available From The Chemical Database Service

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

STRUCTURES

The Structures databases contain a comprehensive collection of organic, organometallic and inorganic compounds, metals,and alloys **crystal structure data**.

CSD - Cambridge Structural Database.
Crystal structure data for over **503,000 organic and organo-metallic compounds**. New releases of this database are received and mounted quarterly.

* **ConQuest** is the graphical front end which has 3D search capabilities.
* **VISTA** performs numerical, statistical and graphical analyses.
* **Mercury** provides comprehensive facilities for visualising crystal structures in three dimensions.
* **IsoStar** A knowledge base of non-bonded interactions derived from the CSD, the Protein Data Bank (PDB) and molecular orbital calculations. Uses a simple web interface.
* **Mogul** A knowledge base of molecular geometries using data derived from the CSD.

ICSD - Inorganic Crystal Structure Data File.
Crystal structure data for **over 124,000 inorganic compounds**

WWW interface available.

CRYSTMET
Crystal structure data for over **126,400 metals, alloys and intermetallics.**

CDIF - Crystal Data Identification File.
Crystal class and unit cell data for **237,671 crystal Structures**.

CrystalWeb
A simple **web interface to all of the crystallographic databases** that allows bibliographic and cell data searching along with structure display.

UTILITY PROGRAMS

A variety of utility programs are available, including links to electronic literature, chemical file format conversion and molecule viewers.

PHYSICAL CHEMISTRY

DETHERM

One of the world's largest thermophysical property databases of pure compounds and compound mixtures. Contains nearly **6 Million data sets** for around **134,000 systems** (around 28,000 pure substances and 106,000 mixtures) covering around **500 property fields**.

Detherm can be accessed via Client server software or via a Web interface.

SPECTROSCOPY

The Spectroscopy databases are designed to aid the chemist in structure elucidation and spectra interpretation problems.

SPECINFO
SpecInfo is a multi-technique spectroscopic database system which covers NMR, IR and mass spectra. A variety of features are available within the program to help with spectrum prediction and searching. It is accessed via the Web interface **SpecSurf**.

ACD/Labs I-Lab
The data can be accessed via **ChemSketch**, a PC based structure drawing package or via the Web using the graphical interface **I-Lab**. In addition to NMR searching and prediction, ACD/Labs I-Lab has compound name generation and property prediction programs.

The databases currently contain:-

No.of Records	13C NMR	1H NMR	15N NMR	17O NMR	31P NMR	19F NMR	11B NMR
SPECINFO	359,647	130,025	7,661	5,184	16,561	25,442	10,238
ACD/Labs	200,100	210,700	9,287	-	27,500	17,000	-

SPECINFO IR Spectra = 20,898 Mass Spectra = 138,727

ACD/Labs Physiochemical data:
pKa = 16,000 LogP = 18,400 Solubility = 5,000

Physical property predictions includes pKa, LogP, LogD, aqueous solubility, boiling point/vapour pressure/enthalpy of vaporization, adsorption coefficient/bioconcentration.

ORGANIC CHEMISTRY

SPRESI

Allows you to search Molecules, Reactions, References and associated data. Contains synthesis planning tool.
Direct access over the web to:-

- 7.0 million structures
- 3.9 million reactions
- 31.7 million factual data entries

extracted from:- 645,000 references
164,000 Patents

Accelrys Accord Databases

Protecting Groups
Selected protection, deprotection and transprotection reactions

Solid Phase Synthesis
Information related to combinatorial synthesis.

BioCatalysis
Biomolecules as catalysts in organic synthesis

Sourcing Databases

CDS Available Chemicals
Up-to-date catalogue data including prices.

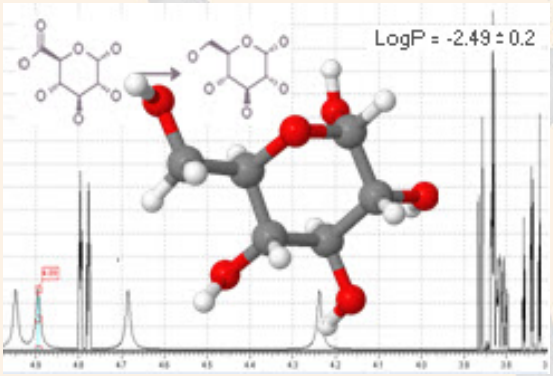
CDS Screening Compounds
Over 3.5 Million entries from 12 Suppliers

General Information

The Chemical Database Service provides on-line access to a variety of quality databases in the field of Chemistry, plus support, training and advice to the UK academic community.

The Service is available **free of charge at point of access** 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 ID's**.

Web Site



Information about CDS, including online help, documentation and up-to-date news items is available over the World Wide Web at the CDS website.

http://cds.dl.ac.uk

Web based interfaces to selected Databases are also available from this site.

Comments:

All comments, questions and suggestions about this newsletter should be sent to:

Dr. D. Parkin
Email: donald.parkin@stfc.ac.uk
Phone: 01925 603 162

Links used in this Newsletter can be found at:
http://cds.dl.ac.uk/letterlinks

Main Service Machine

cds.dl.ac.uk

CDS Helpdesk:
cdsbb@stfc.ac.uk
or phone
01925 603 162

Documentation:

Most documentation is available online or within the client packages. User Guides, training exercises and tutorials can all be downloaded from the CDS web site.

CHEMICAL DATABASE SERVICE
Daresbury Laboratory
Warrington
Cheshire
WA4 4AD
Tel: 01925 603 162
Fax: 01925 603 031
Email: cdsbb@stfc.ac.uk