Crystallographic Databases


CSD


ICSD


MDF

CDIF

What is a database?

- a collection of related data
- tools for adding/amending/updating/records
- tools for selectively extracting data

Crystallographic databases contain

Crystal data

- unit cell dimensions
- space group
- atomic coordinates

- atomic displacement parameters?
- derived information on connectivity?
- data on atom and bond properties?

Crystallographic databases contain

Bibliographic information

author names journal
volume year of publication

Chemical information

formula absolute configuration
systematic name polymorphic form
pharmaceutical/biological activity
experimental details etc.

Crystallographic databases contain

Quality indicators

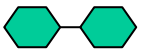
- *R* indices
- the level of precision [e.g., σ(C—C)]
- disorder if present
- any corrections made to original data
- any comments on problems
- any unresolved queries

The exact contents will depend on the database.

Types of search available

... depends on the database

- on a structure fragment (for molecules)



- on a systematic compound name

2,4-dinitro-3-bromophenylhydrazone

Types of search available

- a common or trivial name — acetone, gossypol, Bandrowski's base
- the compound formula — $C_{12}H_{24}AuClS_6$
- element ranges — C_{16} — C_{18}
- compound properties — vasodilator
- experimental conditions — T < 200 K

Types of search available

- space group ($P2_12_12_1$) or number (19)
- unit cell dimensions — *a b c* **a ß ?**
- bibliographic criteria — author, year, etc
- or some combination — OR / AND / NOT

What you need to know

Don't need to know about

the internal database structure

Do need to know about

the various search facilities available

What can you get out?

bibliographic data — atomic coordinates

space group — unit cell dimensions

atomic displacement parameters?

(inter)molecular geometry parameters

analyses of the above — etc.

What are databases for?

- find out what has been done before
- find if specific compounds are known
- check if a unit cell (phase) is known
- get restraints to help structure analysis
- validate and compare your structure

...

What are databases for?

- derive typical geometry parameters
- a source of parameters for calculations
- as a research tool
- find trends/relationships; data mining

What are some limitations?

- unpublished structures not included
- published structures not deposited
- delay in including structures

So check *Chemical Abstracts*, in-house databases, etc.

Crystallographic databases

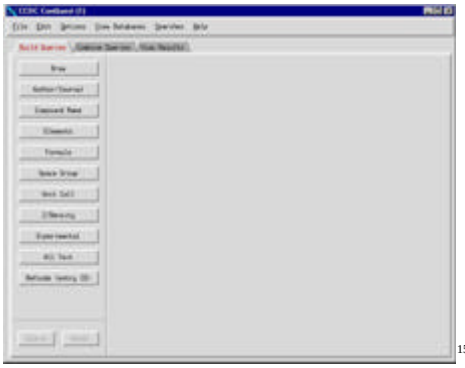
Cambridge Structural Database (**CSD**)

organic and organometallic crystal structures
(not NaCl, PtS, $CuSO_4 \cdot 5H_2O$, metals, alloys;
not proteins or nucleic acids)

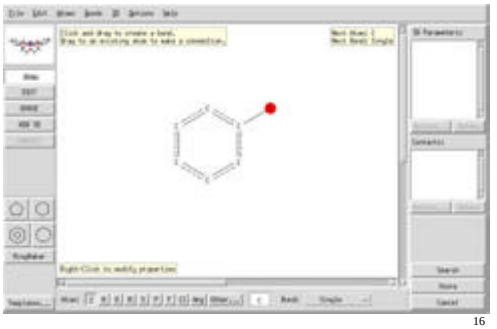
ConQuest/Quest software

- for search, retrieval, display and analysis
- search on the basis of a chemical diagram
- text-based searches also possible

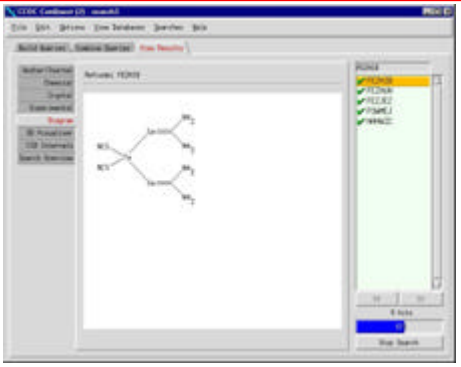
CSD



CSD search fragment



CSD Results



CSD updates

CSD had 335,000 entries by January 2005

full update released every year

+ quarterly Web updates

Related programs

Vista (graphical display of results)

Isostar (library of non-bonded contacts)

Mogul (bonded geometry)

Mercury (graphical display)

Pre-Quest (data input)

Crystal Structure Search and Retrieval (**CSSR**)

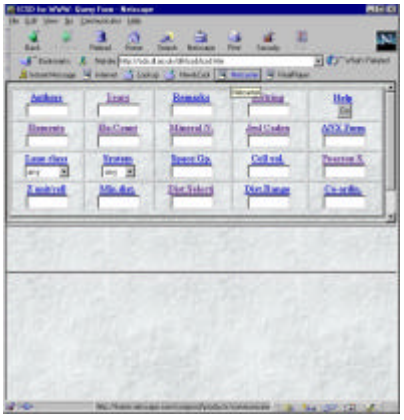
- uses an inverted form of the CSD data
- very efficient at text or formula searches, for example:

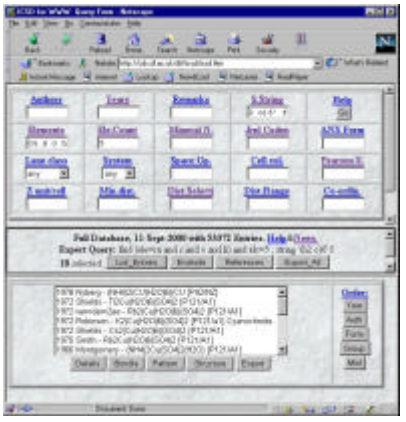
```
AUTHOR Hubberstey
JYEAR  2002
ATOMR  Cu  2  2
INTER  1  2  3
```

- but not for fragment searching

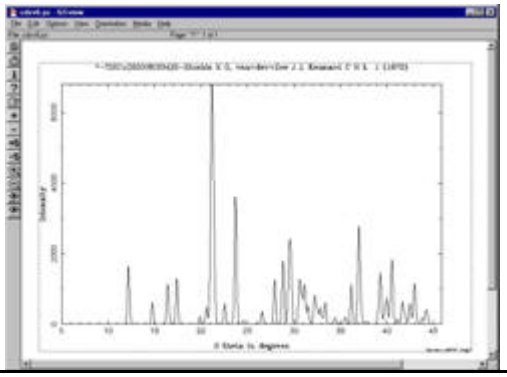
Inorganic Crystal Structure Database (ICSD)

- complementary to the CSD
- non-molecular compounds - in general
- 76,480 entries at April 2004
- command-line interface (like CSSR)
- or Web interface
- no fragment searching (not relevant)
- contains bibliographic information
- contains crystal structure data





ICSD – powder pattern



Metals Data File/CrystMet/MDF

- same types of information as ICSD
- but for metals and alloys
- 75,000 entries at September 2003

Crystal Data Identification File (CDIF)

- unit cell data only
- for over 210,000 entries
- 72,000 do not appear elsewhere
- command-line interface
- superseded by CrystalWeb

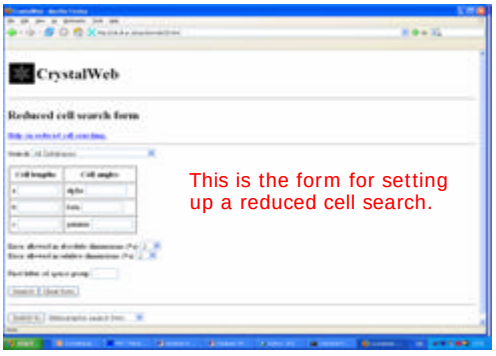
CrystalWeb interface

searches a range of databases:
CSD, ICSD, CrystMet, CDIF

Searches can be:

bibliographic	unit cell
reduced cell	formula
reference code	combined queries

CrystalWeb interface



The Protein DataBank (PDB)

- for biological macromolecules
- bibliographic and co-ordinate details
- about 28,000 structures March 2005
- updated each Wednesday
- UK mirror *http://pdb.ccdc.cam.ac.uk/*

Other databases are available to UK academics via the Chemical Database Service (CDS) at Daresbury Laboratory.