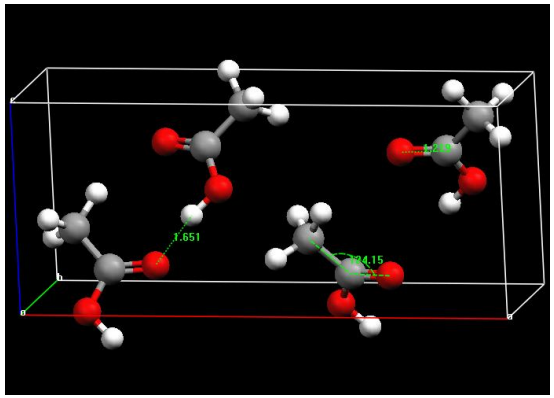
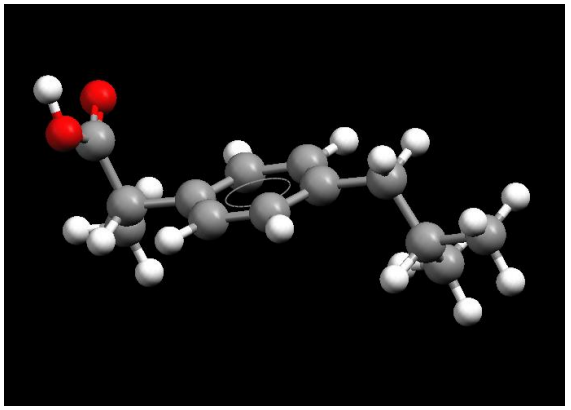
ACS Divison of Chemical Information

Applications of Crystal Structure Information in Chemical Education

Half a million crystal structures in the CSD:

A unique teaching resource in structural chemistry





Frank H. Allen

Emeritus Research Fellow

Cambridge Crystallographic Data Centre, Cambridge, UK

allen@ccdc.cam.ac.uk

Experimental structure determination in chemistry

"There can be no more basic enterprise in chemistry than the determination of the geometric structure of a molecule.

Such a determination, when it is well done, ends all speculation as to the structure, and provides us with the starting point for the understanding of every physical, chemical and biological property of the molecule." ***Roald Hoffman***

Foreword to *Determination of the geometrical structures of free molecules*
[L.V.Vilkov, V.S.Mastryukov, N.I.Sadova, MIR Publishers, Moscow, 1983]

"Crystals are windows on the world of atoms" ***Chet Raymo***

Boston Globe, Science Musings
Collected 'Musings' are in *The Virgin and the Mousetrap*, Viking Books, 1991

Sodium chloride

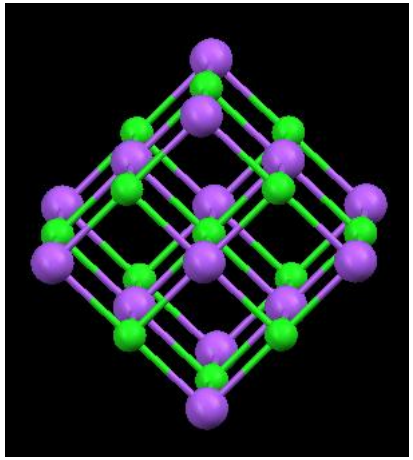
Ferrocene

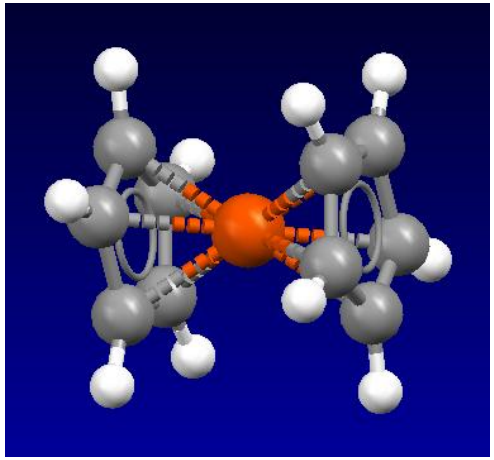
Diamond

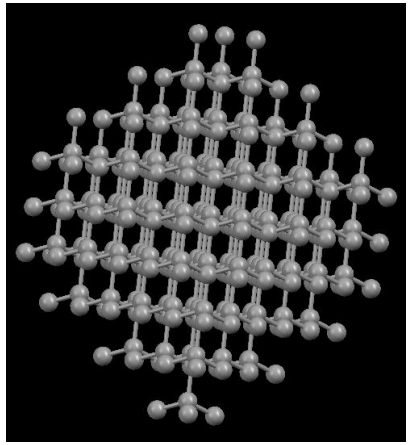
Caffeine

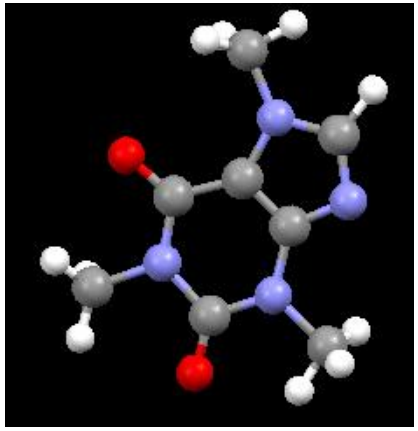
Penicillin

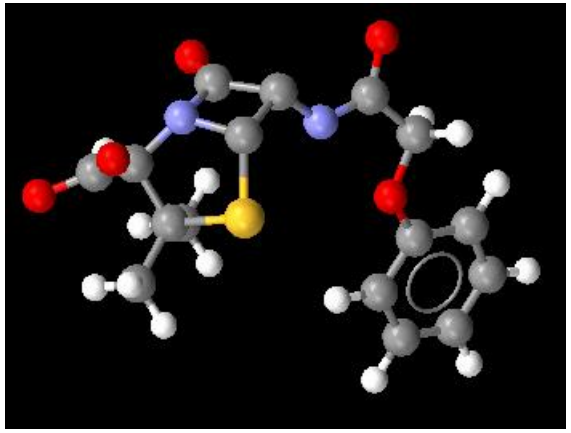
ABP

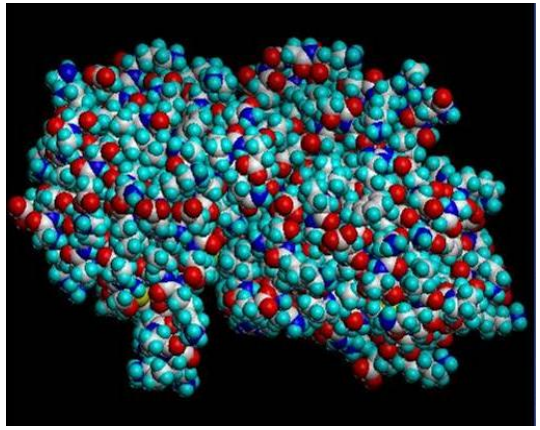












Crystallographic Databases (2009)

Cover the complete chemical spectrum

		Total	*Annual*
CRYSTMET	Metals, alloys, inorganics	119,600	9,000
ICSD	Inorganics & Minerals	100,200	9,000
CSD	**Organics, Metal-Organics**	**488,057**	**40,000**
NDB	Nucleic Acids	3,555	500
PDB	Proteins	50,730	6,000

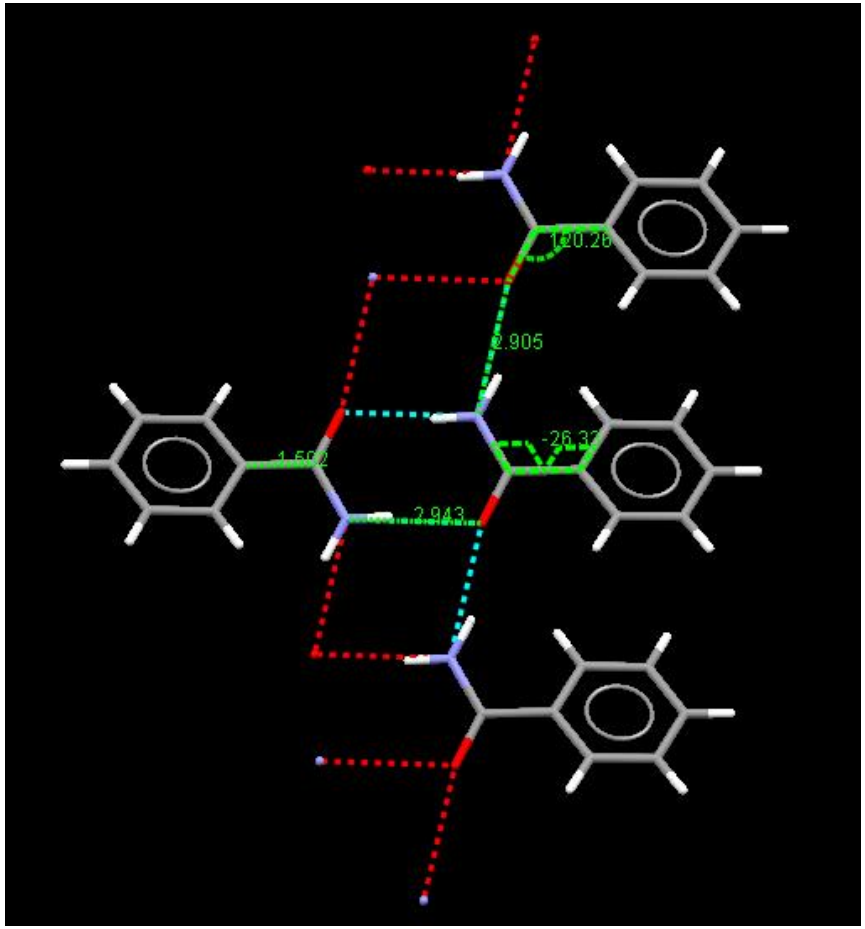
- **The crystallographic databases started in the 1960s & 1970s**
- **If they had not started then, they might never have started at all – start-up costs today would likely be prohibitive**
- **All of these databases have immense value in chemistry teaching**

Small-molecule crystallography in the CSD

Information-rich source of chemical knowledge

Experimental determination of crystal structures at atomic resolution

- **Molecular dimensions**
  Bond lengths, valence angles
- **Molecular shape**
  Torsion angles
- **Intermolecular interactions**
  Distances and angles for H-bonds & other interactions



Scientific Applications of the CSD

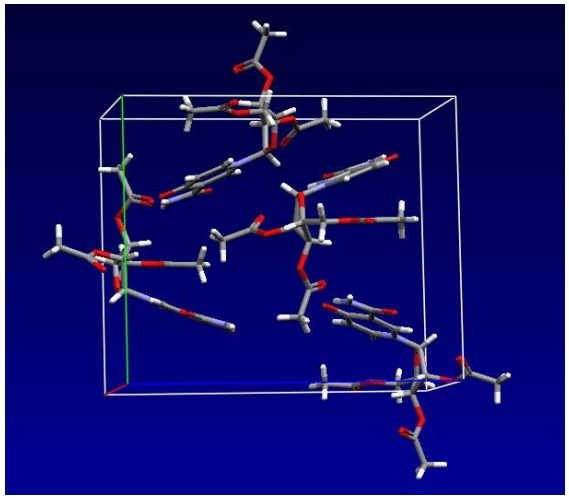
Knowledge (Data) Mining

- **Applied in academia and industry to:**
  - Drug discovery, drug development and formulation
  - Materials discovery and development
  - Structural chemistry
    - Molecular dimensions, Conformational analysis, Reaction pathways, Crystal engineering ……
  - Novel knowledge-based software for solving structural problems
  - Teaching

What we do - The Cambridge Structural Database

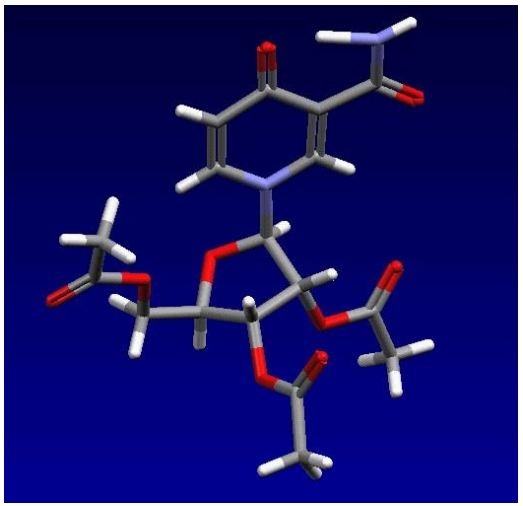
1 - Collect experimental numerical research data

- Repository for the world's output of organic and metal-organic crystal structures (up to ~1,000 atoms), published and unpublished
- Experimental 3D structures from X-ray and neutron diffraction
- Fully retrospective value-added database maintained since 1965
- Collaborative development with researchers and journals worldwide



Crystal Structure

Molecular Structure



Defined experimentally:

- 3D atomic coordinates
- Cell dimensions
- Symmetry

What we do - The Cambridge Structural Database

2 – Curation: Add value to experimental data

- **Bibliographic and chemical text and properties** (all searchable)

4-Oxonicotinamide-1-

(1'-beta-D-2',3',5'-tri-O-acetyl-ribofuranoside)

Source: Rothmannia longiflora

Colour: pale yellow

Habit: acicular

Polymorph: Form IV

C17 H20 N2 O9

G. Bringmann, M. Ochse, K. Wolf,

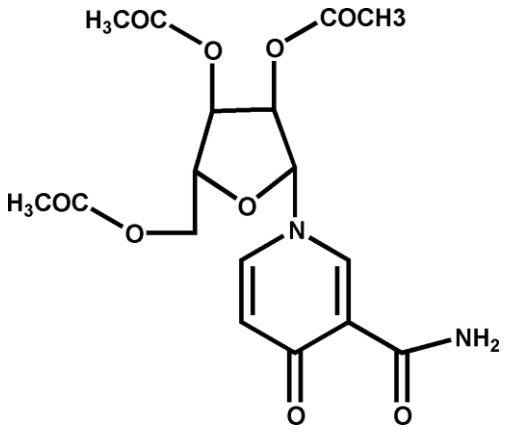
J. Kraus, K. Peters, E-M. Peters,

M. Herderich, L. Ake, F. Tayman

Phytochemistry 51 (1999), p271

R-factor: .0506

- **Chemical diagram and connectivity:**
  Enables 2D & 3D searching for substructures, pharmacophores and intermolecular interactions



- **DOI link to original paper**

What we do - The Cambridge Structural Database

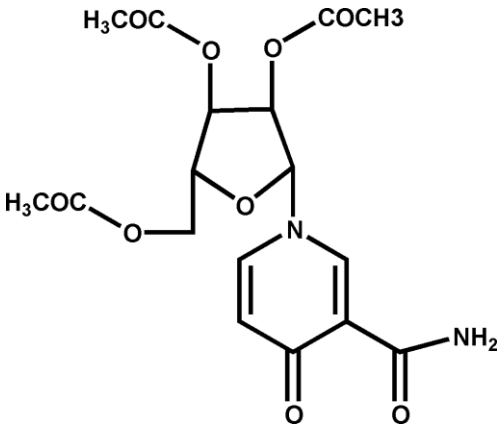
3 – Validate information

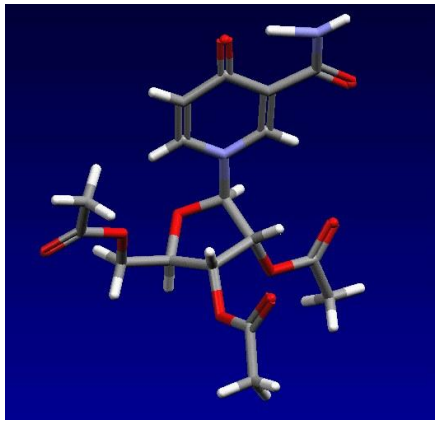
Molecular structure

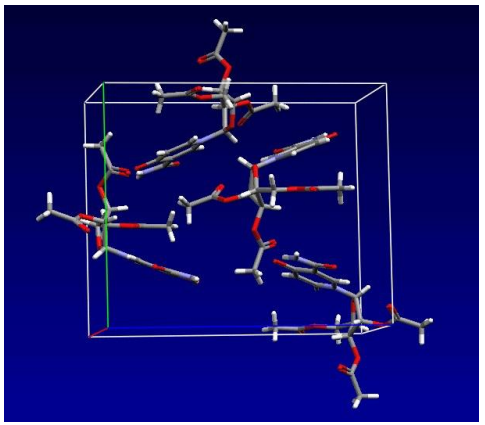
- Structure must match name and formula
- Chemical and crystal structures must match
- Calculated geometry of molecule must match published geometry and be in normal ranges

Crystal structure

- Unit cell & symmetry checks
- Acceptable intermolecular contact distances
- Analysis of crystallographic disorder

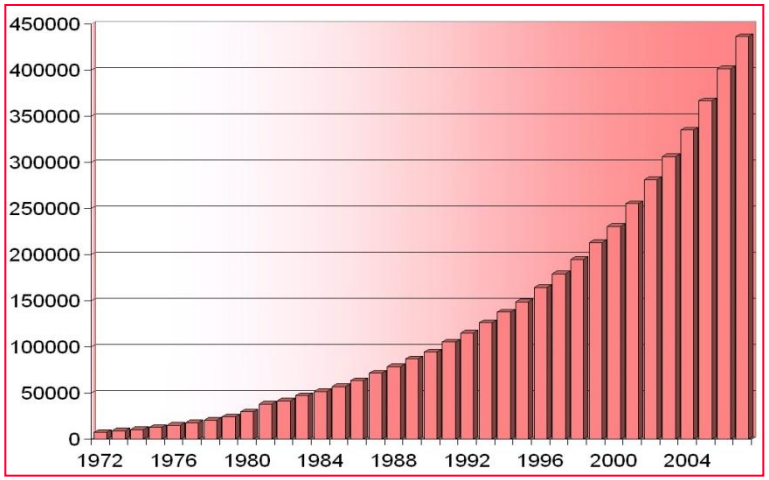






How much data have we accumulated?

"All science is either physics or stamp collecting" (Lord Rutherford)



Growth 1970 – 2008:

488,057 structures

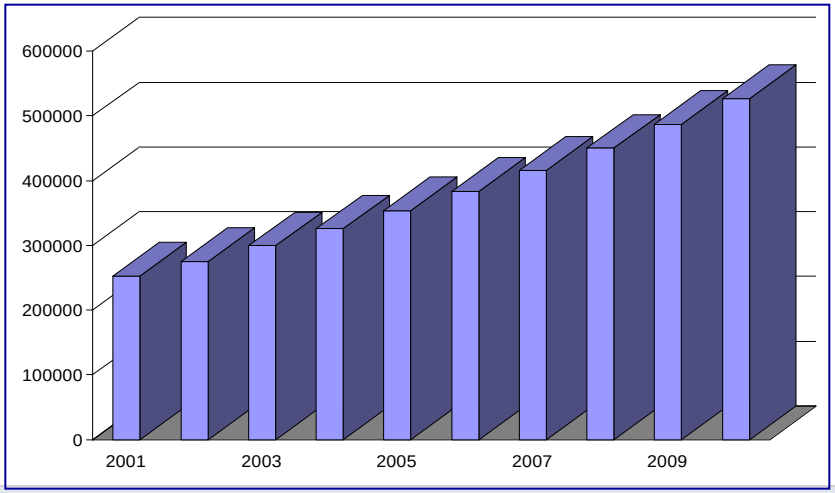
on

15 July 2009

Projected Growth 2001-2010:

>500,000 structures

by end of 2009



CSD records coordinates for 35,144,049 atoms

CSD: Overview of structure types

July 2009 database, 488,057 structures

- **Organic (209,148 = 42.9%)**

Amino ac/peptides	5,848	Carbohydrates	5,574
Terpenes	4,426	Steroids	3,773
Alkaloids	2,650	Nucleosides/tides	1,909

- **Metal-organic (257,206 = 52.7%)**

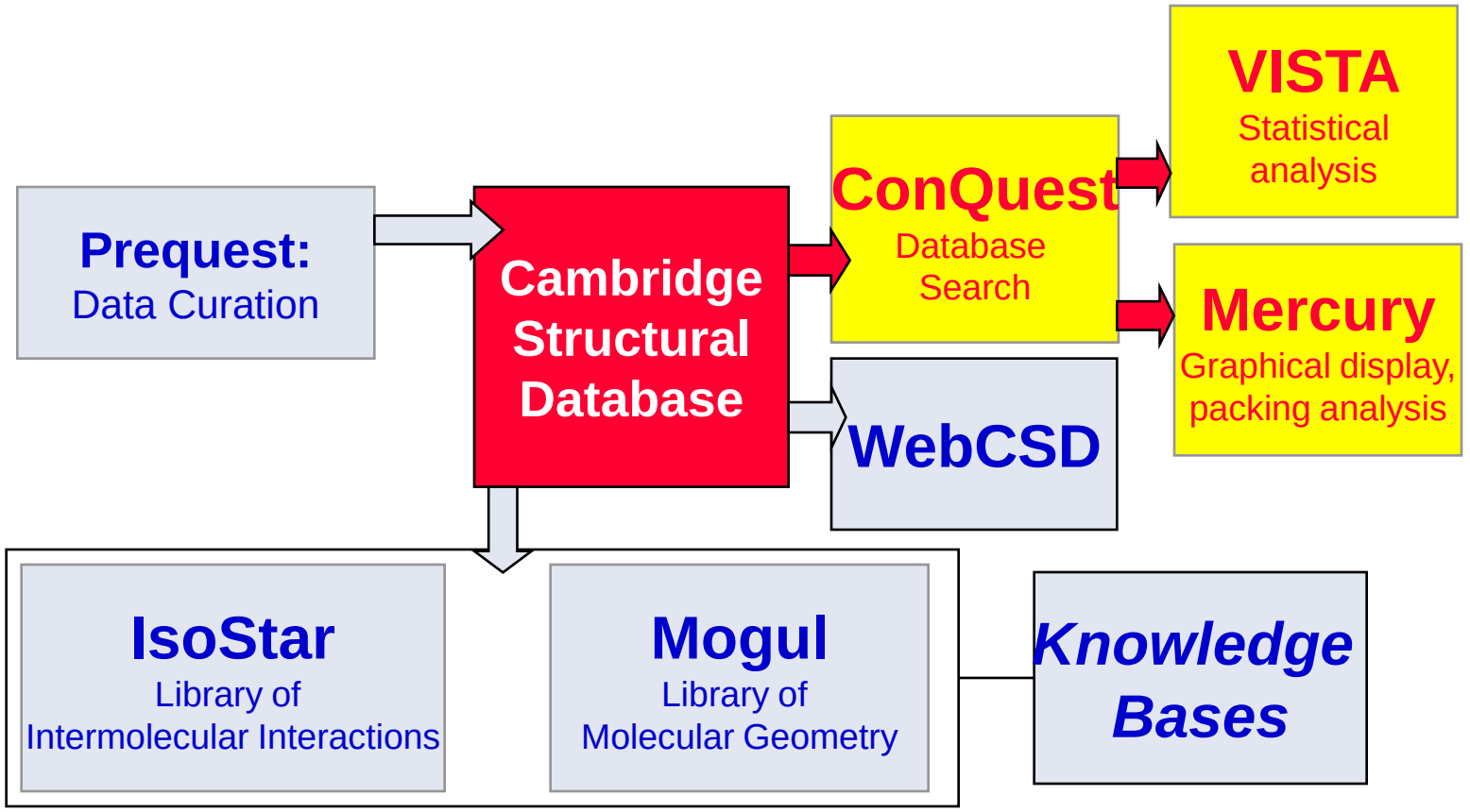
Fe	25,916	Cu	30,356	Pt	12,136
Os	4,676	Hf	712	Ln/An	16,155

- **Main Group Cpds (31,235 = 6.4%)**

Se	6,837	Sb	4,281	Xe	22

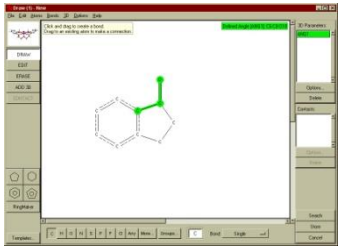
The Cambridge Structural Database System

System components



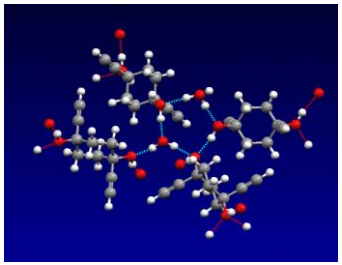
The Cambridge Structural Database System

Software Applications



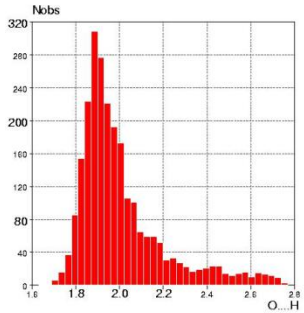
ConQuest

- **Search** all CSD data including 2D and 3D substructures and intermolecular interactions
- Browse hits
- Retrieve CSD information, including geometry



Mercury

- **Visualise** molecules, crystal packing, H-bonded networks, slices through crystals, etc.
- Locate intermolecular interaction motifs
- Wide variety of display styles



Vista

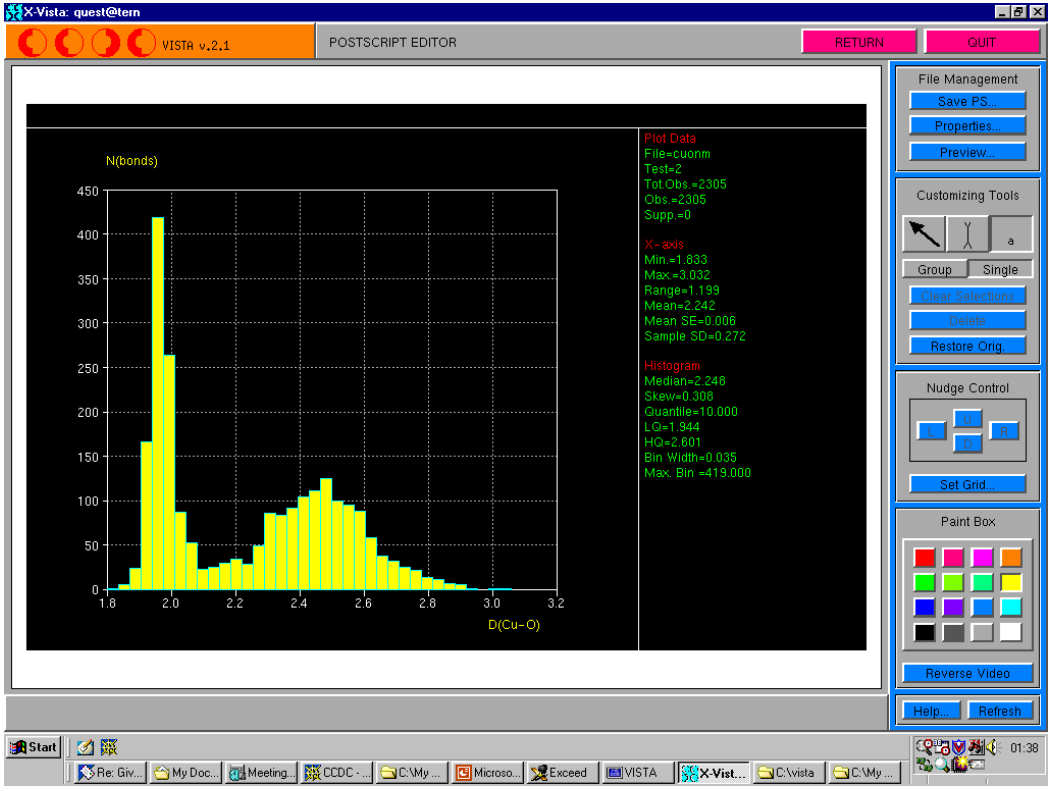
- **Analyse** and display retrieved data
- Plot histograms, scattergrams, polar plots
- Perform statistical operations (PCA etc.)

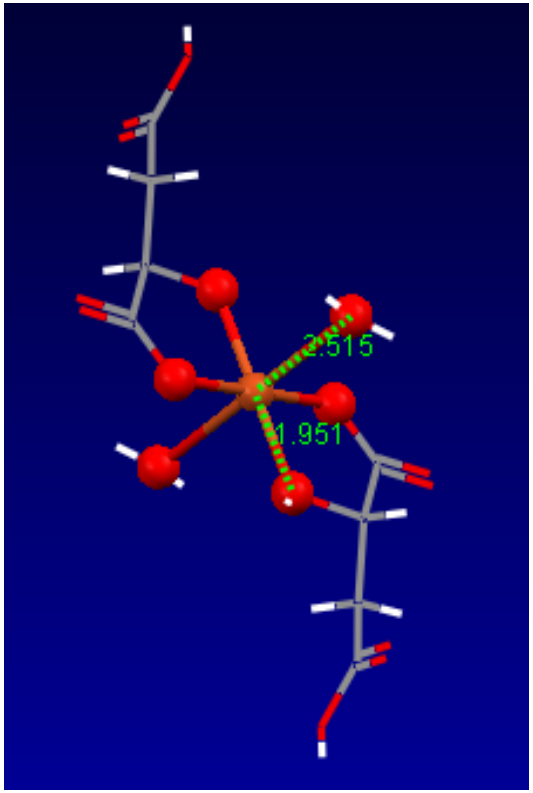
What does the CSD offer to chemistry teachers?

- Search, view and manipulate 0.5 million precise 3D structures, including chemical bond types and geometry
- Real experimental data, error estimates, variance, etc.
- Exemplify and quantify:
  - Molecular shape, stereochemistry, metal coordination spheres, etc.
  - Hydrogen bonding, molecular recognition and supramolecular chemistry
- Massive structural diversity in organic and metal-organic chemistry
- **Common molecules abound, unusual molecules are common!**

Discovering chemistry using the CSD

Bond length variations: Jahn – Teller effect



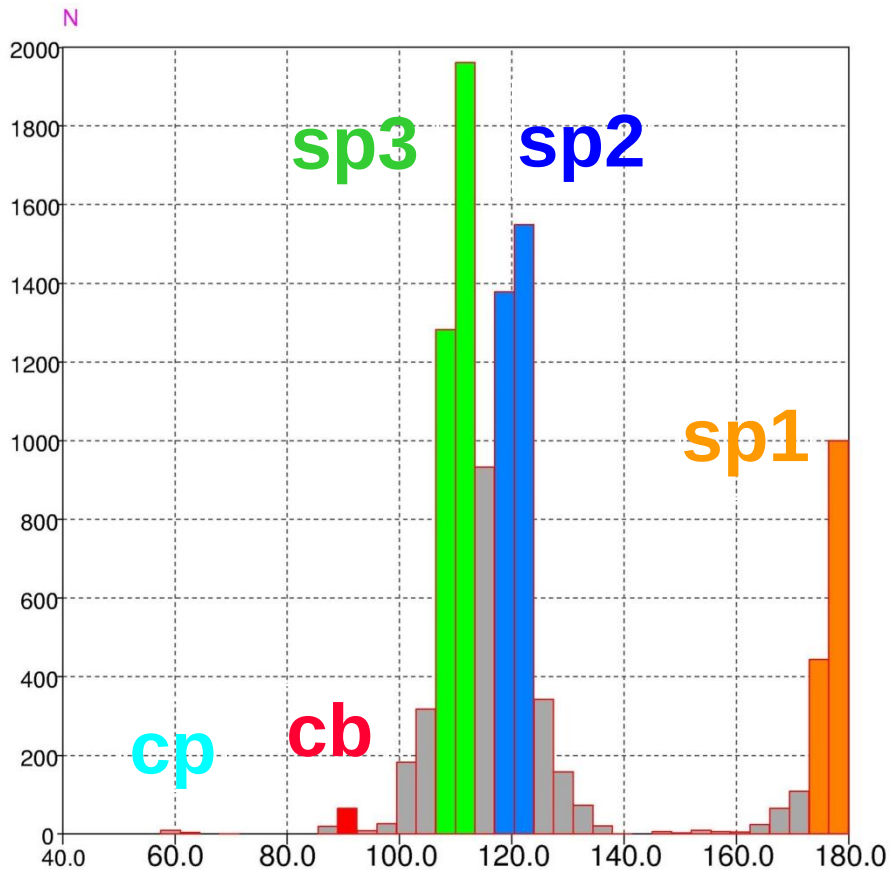


ConQuest search for CuO_6 octahedra, retrieve Cu-O distances, plot in VISTA

Examine individual structures in Mercury

Discovering Chemistry using the CSD

‘Spectrum’ of C~C~C valence angles (~ any bond type) (VISTA)



Can initiate discussion of:

Hybridisation of C and its effect on valence angles

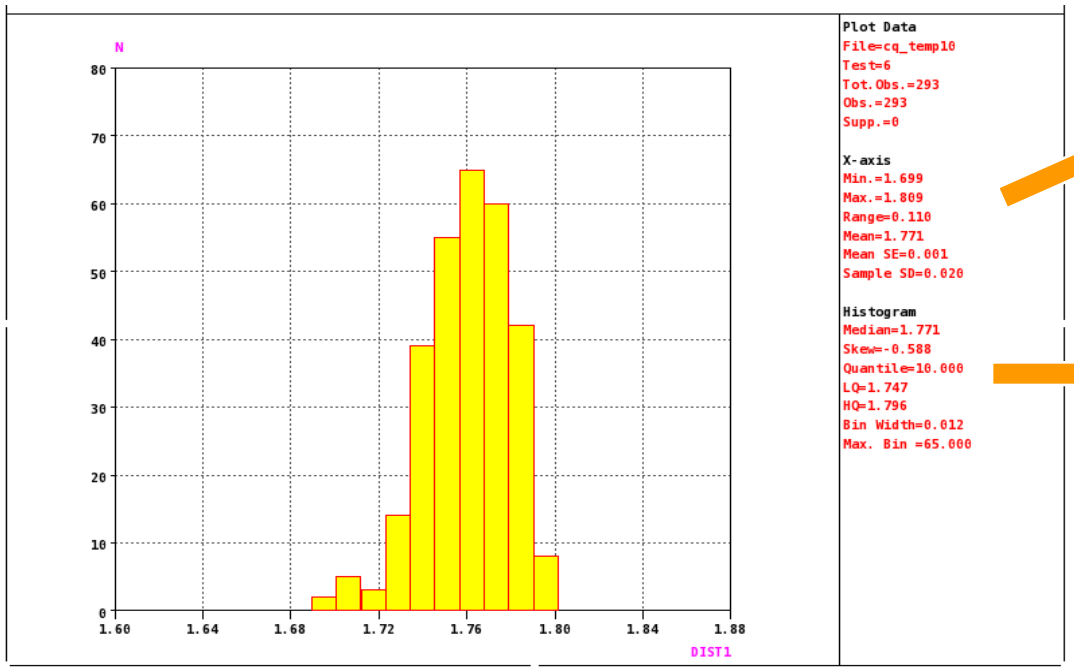
Strain in small rings – Cyclopropane and cyclobutane

Relaxation of strain at normal C(sp3) atoms – cyclic or acyclic

Discovering chemistry using the CSD

Exploring real experimental data (VISTA)

Vista plot of C(cyclobutyl) – Cl bond lengths: Variation of experimentally determined values and simple descriptive statistics

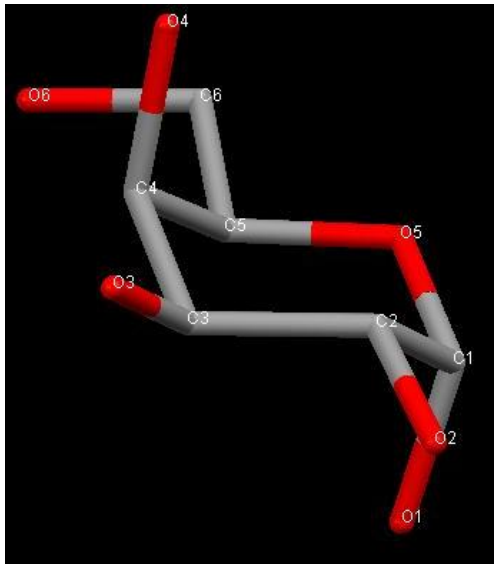


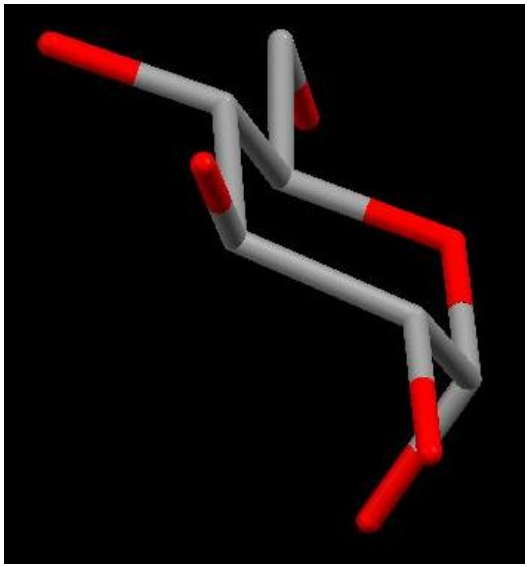
X-axis
Min.=1.699
Max.=1.809
Range=0.110
Mean=1.771
Mean SE=0.001
Sample SD=0.020

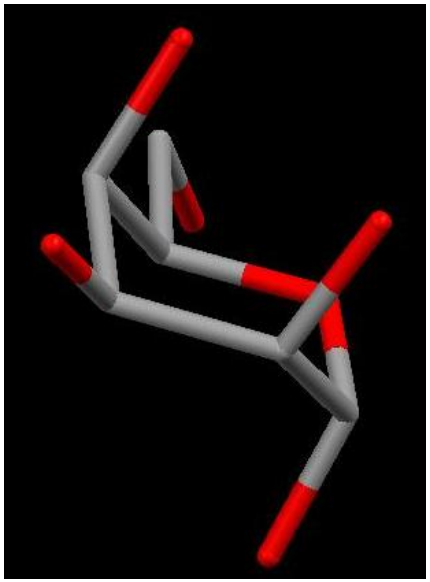
Histogram
Median=1.771
Skew=-0.588
Quantile=10.000
LQ=1.747
HQ=1.796
Bin Width=0.012
Max. Bin =65.000

Learning about 3D conformations

Stereochemistry of α-D-glucopyranoses (Mercury)



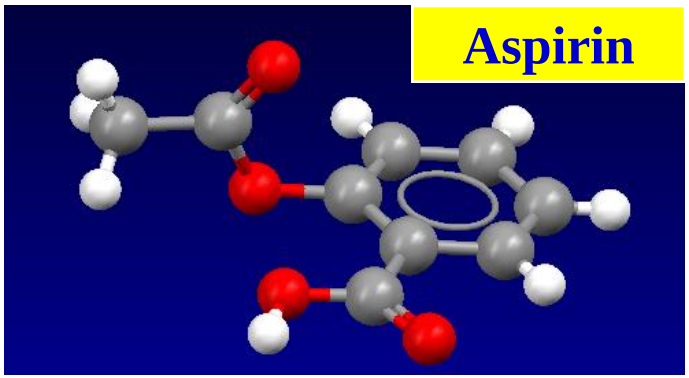




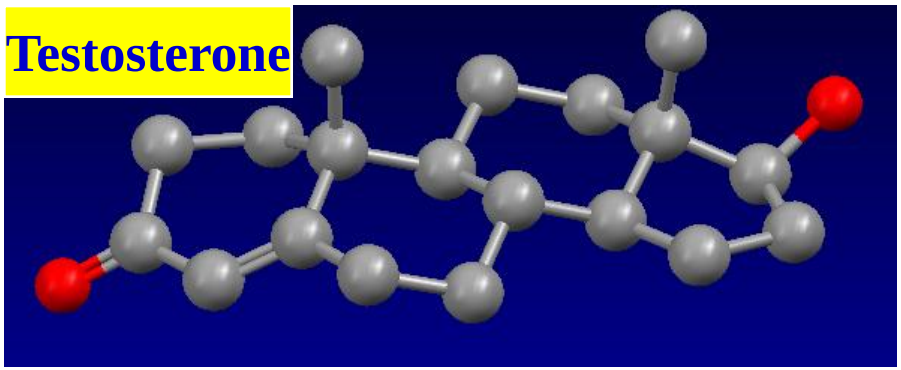
α-D-galactose	α-D-glucose	α-D-talose
a, e, e, a, e	a, e, e, e, e	a, a, e, a, e

Learning about 3D chemistry

Some important organic molecules (Mercury)

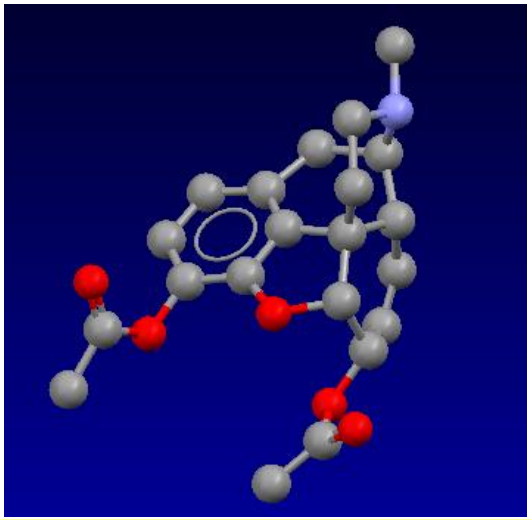


Aspirin

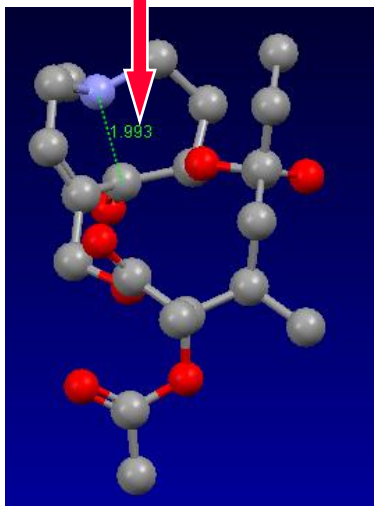


Testosterone

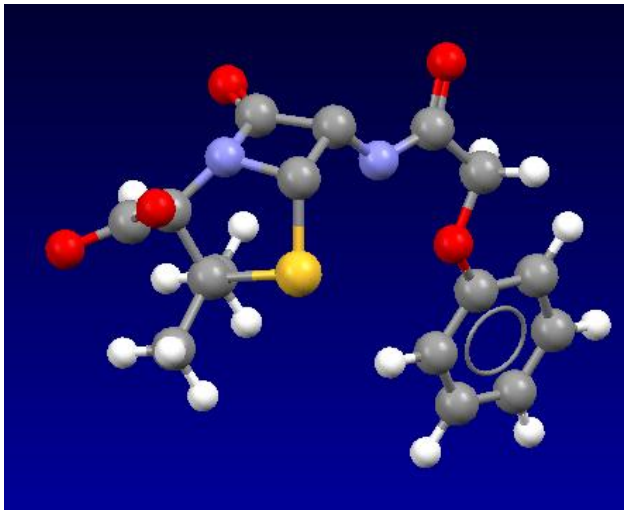
1.993



Diacetylmorphine (heroin)



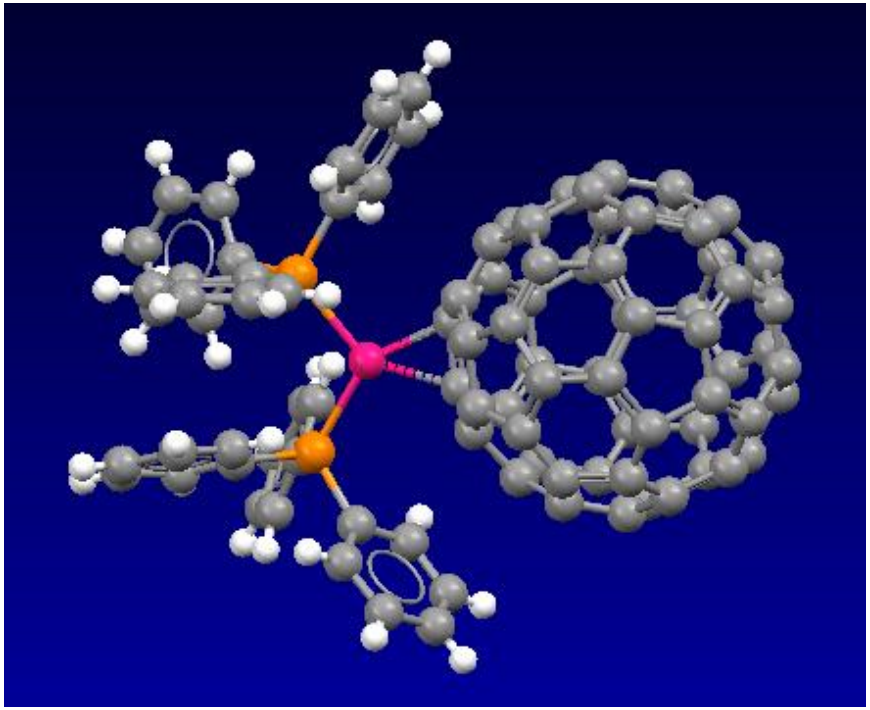
Clivorine



Phenoxymethylpenicillin

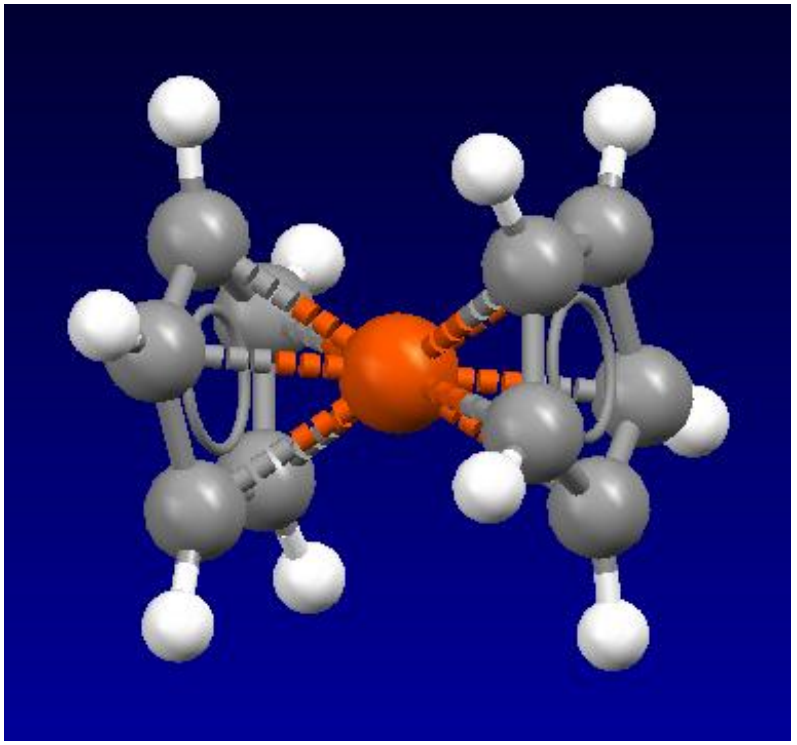
Learning about 3D chemistry

Important inorganic molecules (Mercury)



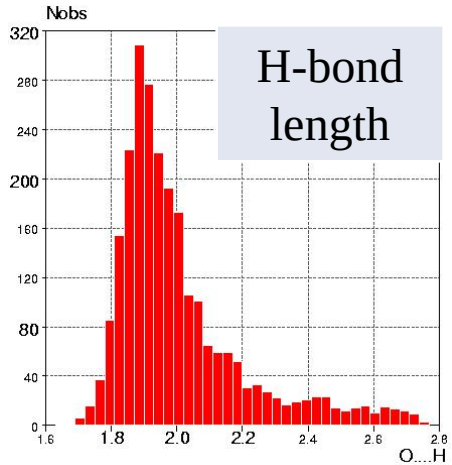
$Pt(tpp)_2$ complex of C60-fullerene

Ferrocene

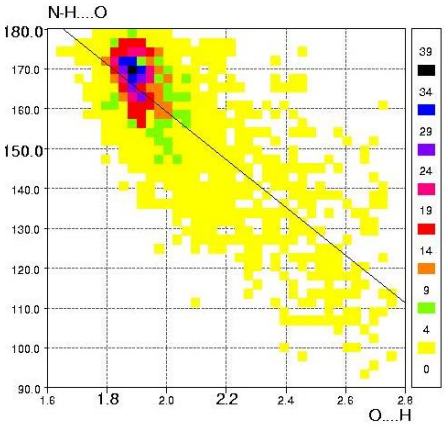


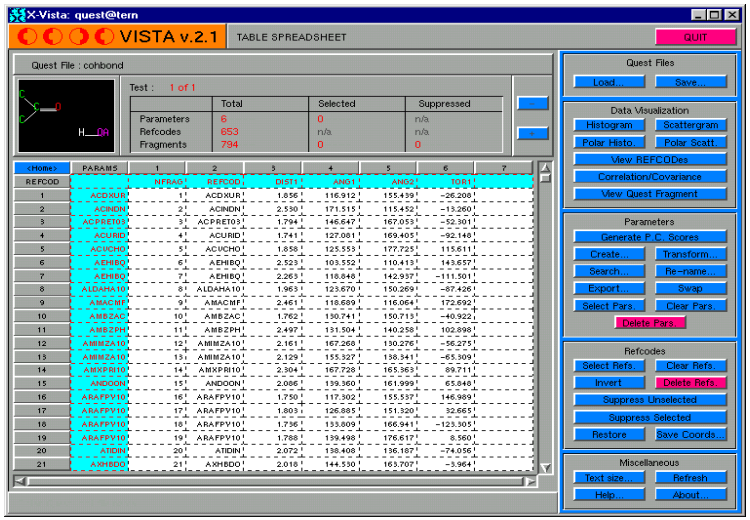
CSD System Software

N-H…O (amide) hydrogen bonding (VISTA)

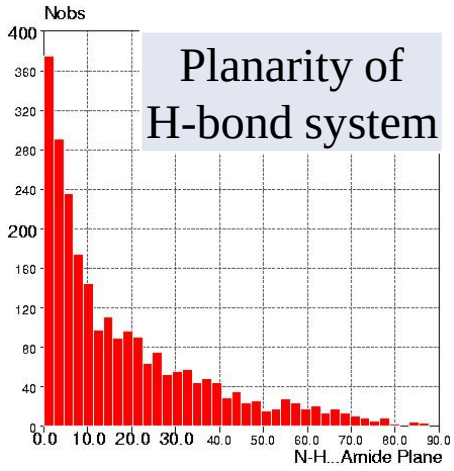


H-bond length

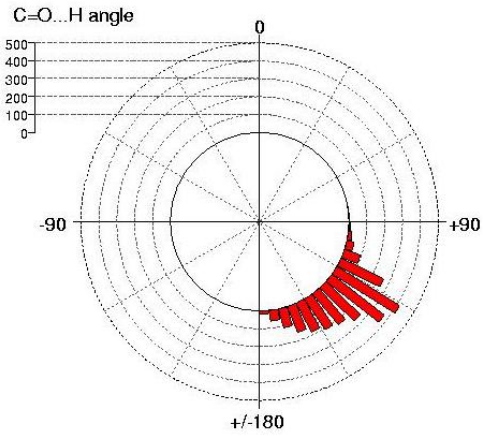




Spreadsheet of H-bond geometry



Planarity of H-bond system

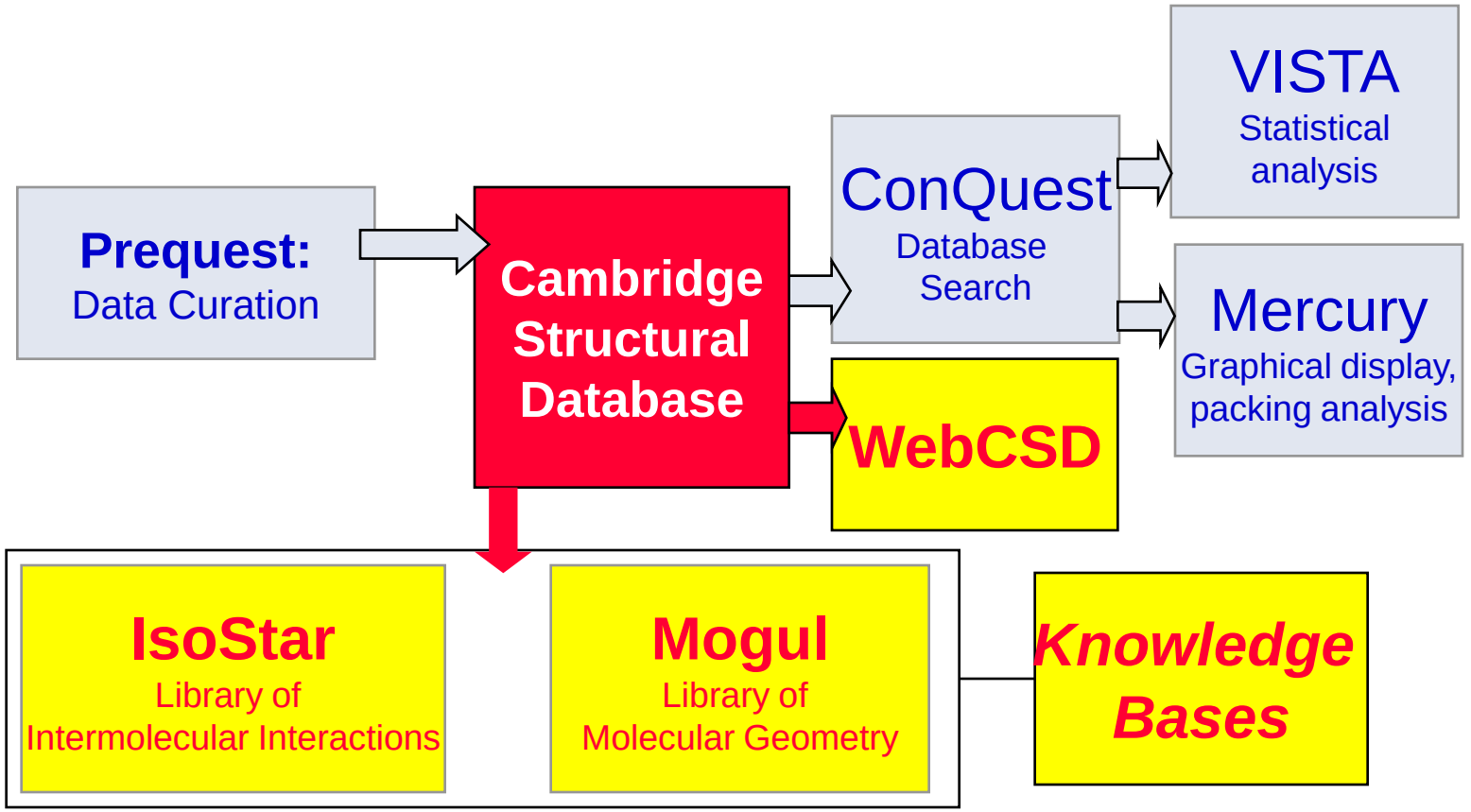


Directionality of H-bond

at H ← → at O

The Cambridge Structural Database System

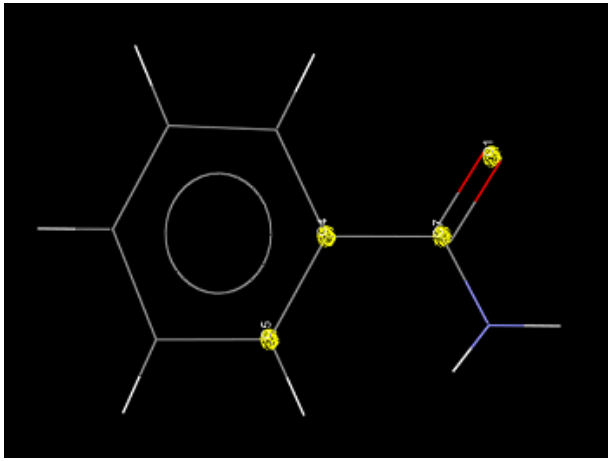
Knowledge-based libraries and WebCSD



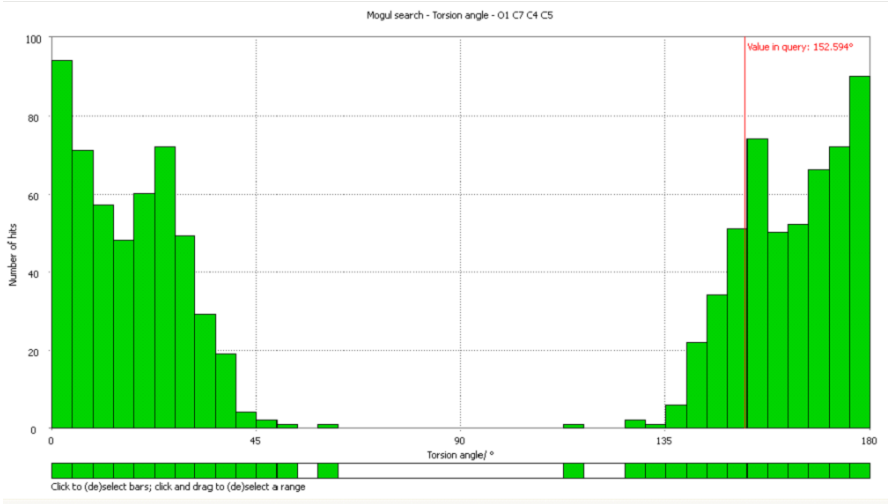
Mogul –A Library of intramolecular Geometry

Rapid access to bonds angles and torsions

- 20 million bond lengths, valence angles and torsions, 1.8 million parameter distributions
- Import or draw molecule into Mogul interface
- Click on 2, 3 or 4 atoms (bond, angle, torsion)
- Search chemical attribute tree and display parameter distribution



Torsion angle

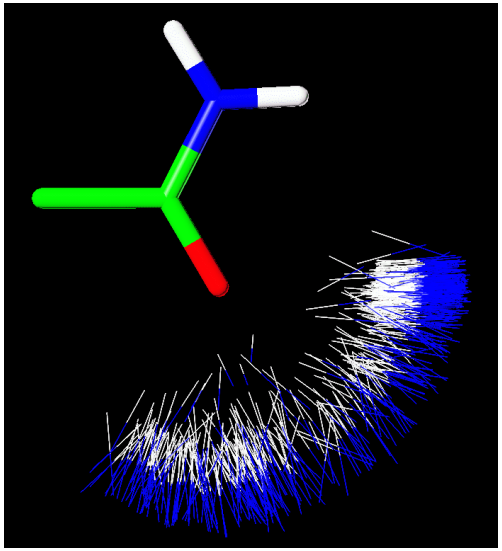


Bruno et al., JCICS, 44, 2133 - 2144, 2004

IsoStar Library of Intermolecular Interactions

Scatterplots and contoured plots

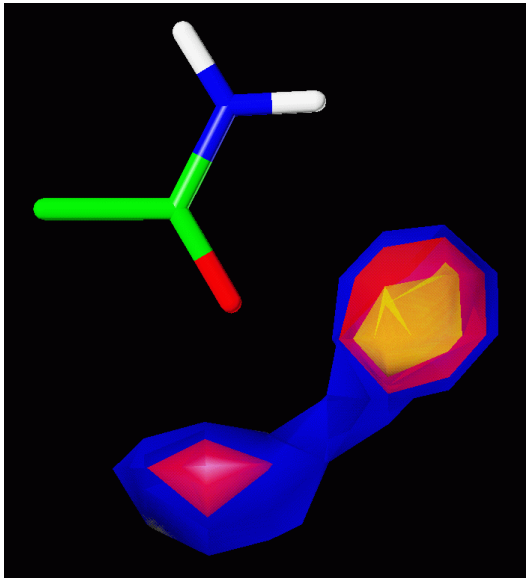
- Data from CSD, Protein Data Bank and computed interaction energy minima
- Covers 300 central groups and 48 contact groups
- Interaction distributions displayed as scatterplots or contoured plots
- Contains 20,000 CSD scatterplots, 5,500 PDB plots and 1,500 energy minima



Distribution of N-H donors around amide C=O group

Basic plot

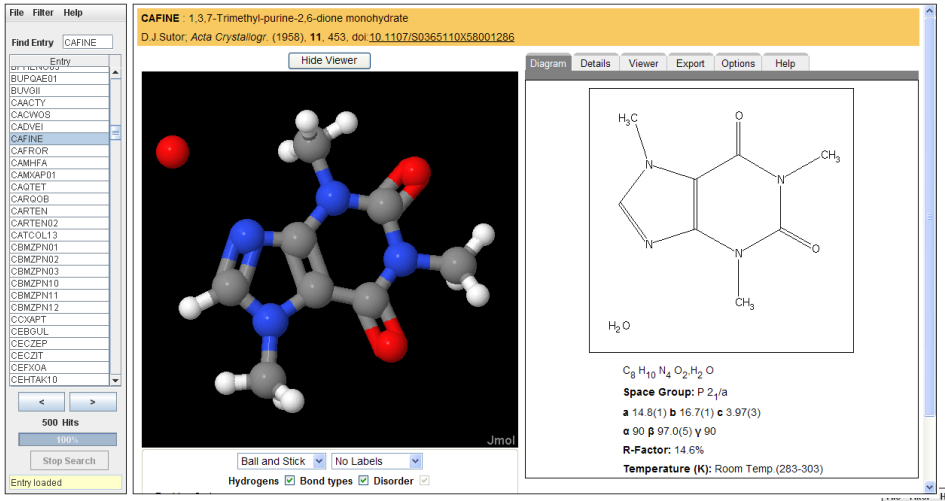
Contoured plot



Bruno et al., J.Comp.-Aided Mol. Des., 11, 525-537, 1997

WebCSD (released in July 2009)

CSD on the Internet for researchers and teachers

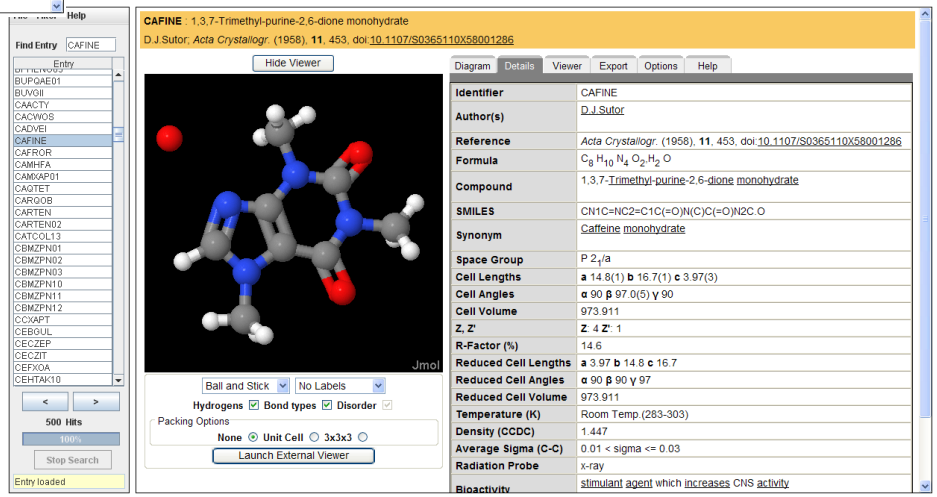


Available since July 2009

Active further development including intranet version

Full display of CSD entry information

- **2D substructure searches**
- **Chemical similarity searches**
- **Text and numeric searches**
- **3D search and geometry retrieval now being added**



History of the CCDC

- **Founded in 1965, Dept. of Chemistry at Cambridge**
  - Grant funding from UK agencies
- **Independent Institution since 1987**
  - Not-for-profit Company with charitable status
  - International Board of Trustees
  - Close links with University of Cambridge, inc. PhD supervision
  - 50 staff in 2009 – inc.15 CSD Staff, 12 software engineers
- **CSD System is a cost-recovery subscription service**
  - 130 Industrial company sites
  - ~ 1,400 Academic sites in 67 countries

Free teaching materials on the CCDC website

Go to www/ccdc.cam.ac.uk/free_services/teaching

- **Ten CSD teaching examples**
- Subset of **500 CSD entries** linked to teaching examples
- Use **WebCSD** to access and examine subset and work through teaching examples
- Structures can also be visualised using the free version of **Mercury** (also downloadable from CCDC website)

- Teaching tools will be further discussed in this Symposium by Gary Battle and Greg Ferrence

Acknowledgements

CCDC Staff, May 2008 plus
Greg Ferrence (ILSTU, inset left) and Colin Groom (Director CCDC, inset right)