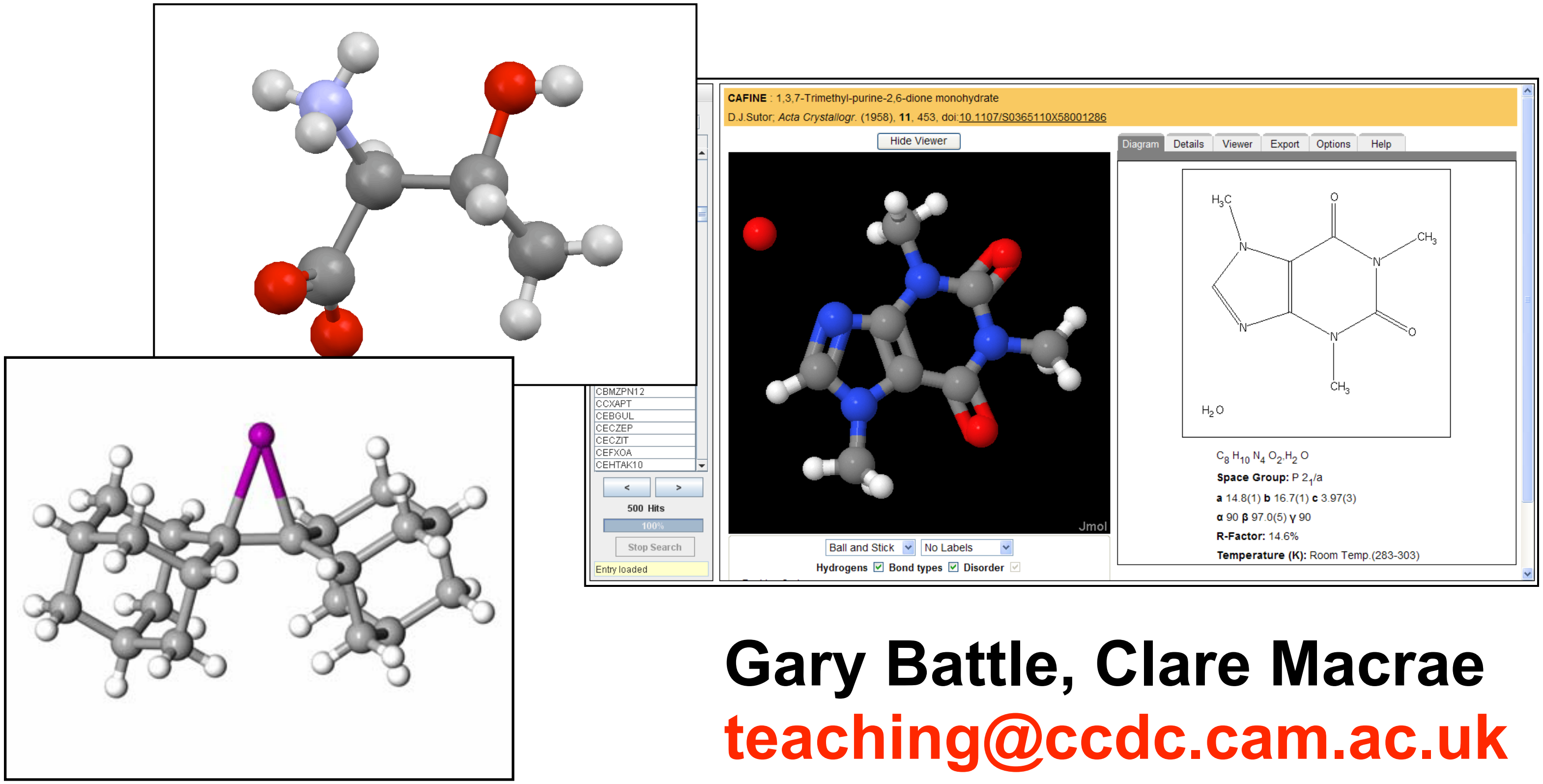
Imperial College Cambridge Structural Database Workshop



Gary Battle, Clare Macrae
teaching@ccdc.cam.ac.uk

Schedule…

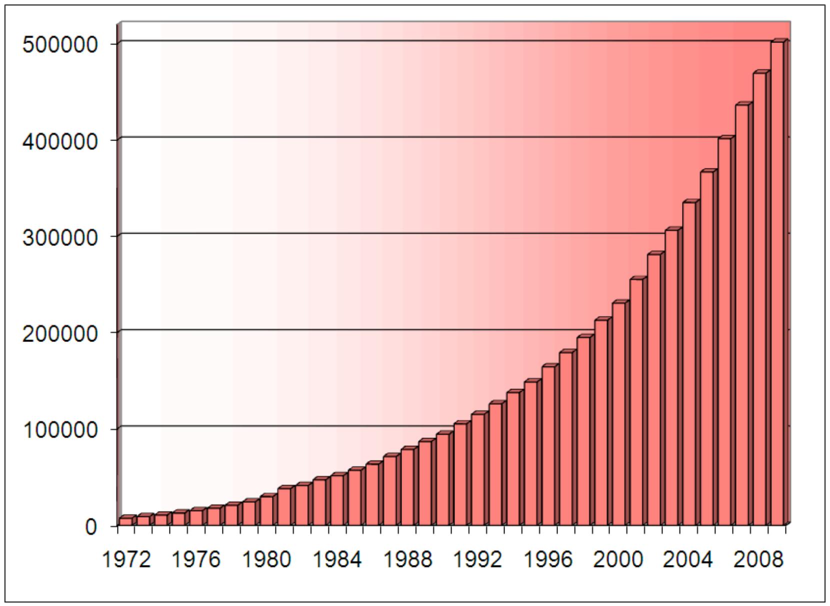
- Introduction: CSDS content and coverage, use in teaching structural chemistry
- Web-based access: software tools and example uses
- Advanced desktop applications: software tools and examples
- IsoStar library of intermolecular interactions: overview and demo
- SuperStar: predicting interactions

Crystal Structures are Uniquely Suited for Teaching Structural Chemistry

- X-ray analysis method of choice for 3D structure characterization
- Very precise: standard uncertainties <0.005 Å on bond lengths and <0.5 on valence and torsion angles
- Provide remarkable richness of structural information: both the 3D geometric structures of molecules and also the nature and geometry of their interactions
- > 750,000 crystal structures published in past 60 years, and modern instrumentation adding > 60,000 novel structures each year

Cambridge Structural Database

Worldwide repository of validated small-molecule organic & metal-organic crystal structures



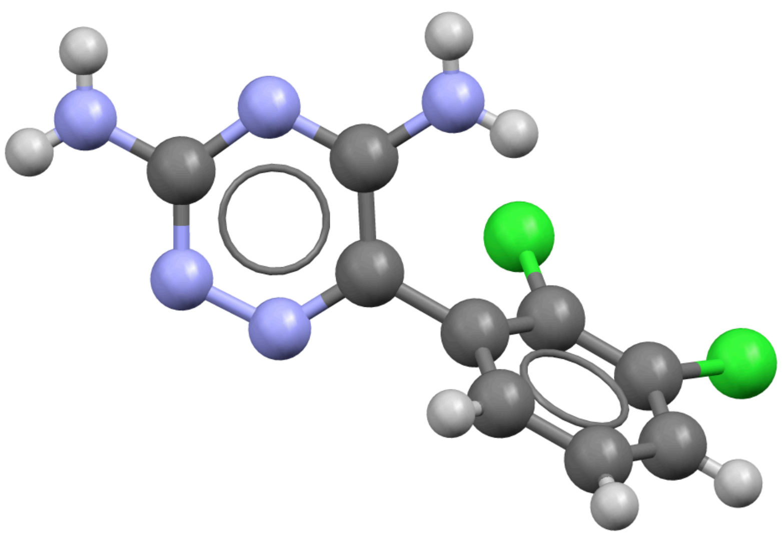
CSD Growth 1970-2010

Dec 09 – 500,000th structure milestone reached

Lamotrigine

Acta Cryst., Sect.C:Cryst Struct. Commun. (2009), **65**, o460

Refcode: **EFEMUX01**



Pedagogical Value

- 3D visualisations enhance students conceptual understanding and spatial abilities

  - Williamson, V. M. *J. Chem. Educ*., **2008**, 85, 718-723
  - Bodner, G. M.; Guay, R. B. "The Purdue Visualization of Rotations Test" *The Chemical Educator*

- Use of experimentally measured data is of great pedagogical value, and has been shown to enhance student learning

  - DeHaan, R. L. *J. Sci. Educ. Technol*., **2005**, 14, 253-269
  - Handelsman, J. *Science* **2004**, 304, 521-522
  - Prince, M. *J. Eng. Educ.,* **2004**, 89, 1-9

Information • Textbooks • Media • Resources

Using the Cambridge Structural Database to Introduce Important Inorganic Concepts

Tiana V. Davis, M. Shahzad Zaveer, and Marc Zimmer*
Chemistry Department, Connecticut College, New London, CT 06320; **mzim@conncoll.edu*

Data and structure correlation analysis is an increasingly important area in science *(1)*. This is particularly true in biology where genomic and proteonomic studies are generating vast amounts of data. In order to expose our students to chemoinformatics and introduce them to an important resource in chemistry, the Cambridge Structural Database, we have devised a simple series of inorganic exercises that can be done in introductory inorganic classes.

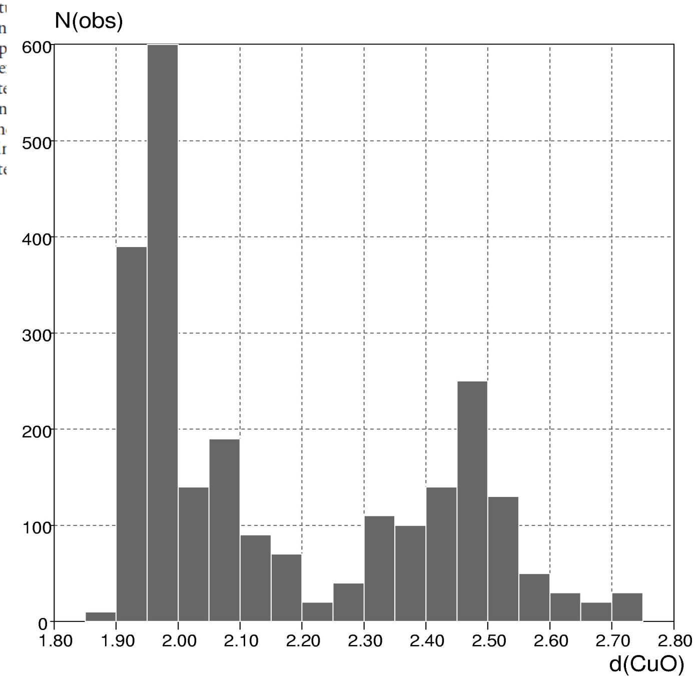
The latest version of the Cambridge Structural Database (CSD), version 5.22, contains the CSD database with 245,392 structures, ConQuest (an interface to CSD), Mercury (a visualization program), Isostar (software for superimposing molecular fragments) and Vista (a statistics package designed for use with the CSD). CSD has 245,392 X-ray and neutron diffraction structures of organocarbon compounds *(2, 3)*.[1] All the compounds in the CSD have less than 1,000 atoms. Peptides with up to 24 residues are covered *(4)*, while larger peptides and proteins are in the Protein Database.[2] The CSD has 109,349 structures that contain one or more transition metal ions. Classroom ConQuest comes with a reduced

of congeneric families are useful as they can reveal the different backbone conformations the structure can adopt in the different crystals environments. These analyses, in turn, can provide information about conformations available to the backbone, interconversions of the conformers, and environmental factors responsible for certain conformations *(6)*.

The release of a new user-friendly graphics interface called ConQuest, easy to use tutorials, a Windows PC and Linux version, and a classroom edition have prompted us to develop some inorganic laboratory exercises utilizing the database. The purpose of these exercises is to expose the stu dents to database analyses, and to demonstrate inorgan structural properties and structure correlation. Four examp exercises are outlined below; all the structures used in the e ercises are shown in Table 1. The exercises can be presente as open-ended discovery assignments or as more tradition problems. As the tutorials are complete, easy to understan and instructive, the students need no assistance in doi ConQuest searches. The resultant data can easily be exporte to Isostar, Vista, or Excel.

J. Chem. Educ., **2002**, 79 (10), p 1278,
DOI :10.1021/ed079p1278

Backbonding
Eighteen-Electron Rule
High-Spin vs. Low-Spin
Jahn-Teller Effect



In the Classroom

Teaching with Technology

edited by
Gabriela C. Weaver
Purdue University
West Lafayette, IN 47907

Using the Cambridge Structural Database To Teach Molecular Geometry Concepts in Organic Chemistry

Jay Wm. Wackerly, Philip A. Janowicz, Joshua A. Ritchey, Mary M. Caruso, and Erin L. Elliott
Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Jeffrey S. Moore*
Departments of Chemistry, Materials Science and Engineering, and the Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801; **jsmoore@illinois.edu*

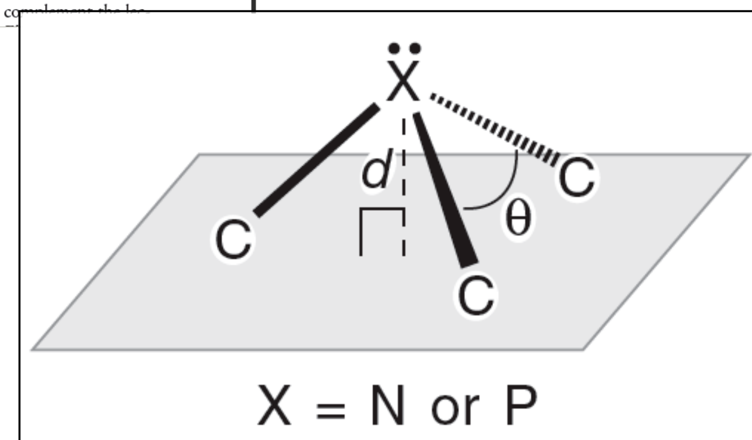
A large lecture classroom may not seem to give students a chance to move beyond their textbook structures and analyze data from actual molecular structures. However with continuous technological advances applied to education, individualized software programs are more commonly available to enhance students' learning *(1)*. It has been shown that students take a greater interest in their coursework when they have the opportunity to analyze "real-world" data *(2)*. Given society's growing dependence on information technology and the need to understand the basic concepts of data mining and parameter correlation, the opportunity to introduce these ideas in an organic chemistry classroom full of future professionals is appealing *(3)*. Herein

The CSD is a database of small organic and organometallic structures that have been elucidated by X-ray or neutron diffraction techniques *(5)*. The database currently has 400,977 structures, and 43% of these are organic compounds. In order to give all of the students convenient access to the database, Classroom ConQuest was used.[1] Classroom ConQuest is provided by the Cambridge Crystallographic Data Center free of charge[2] although it contains a reduced number of structures (11,300). While only a subset of the full CSD, meaningful searches using Classroom ConQuest for simple structural trends can still be obtained.

These assignments were designed to co

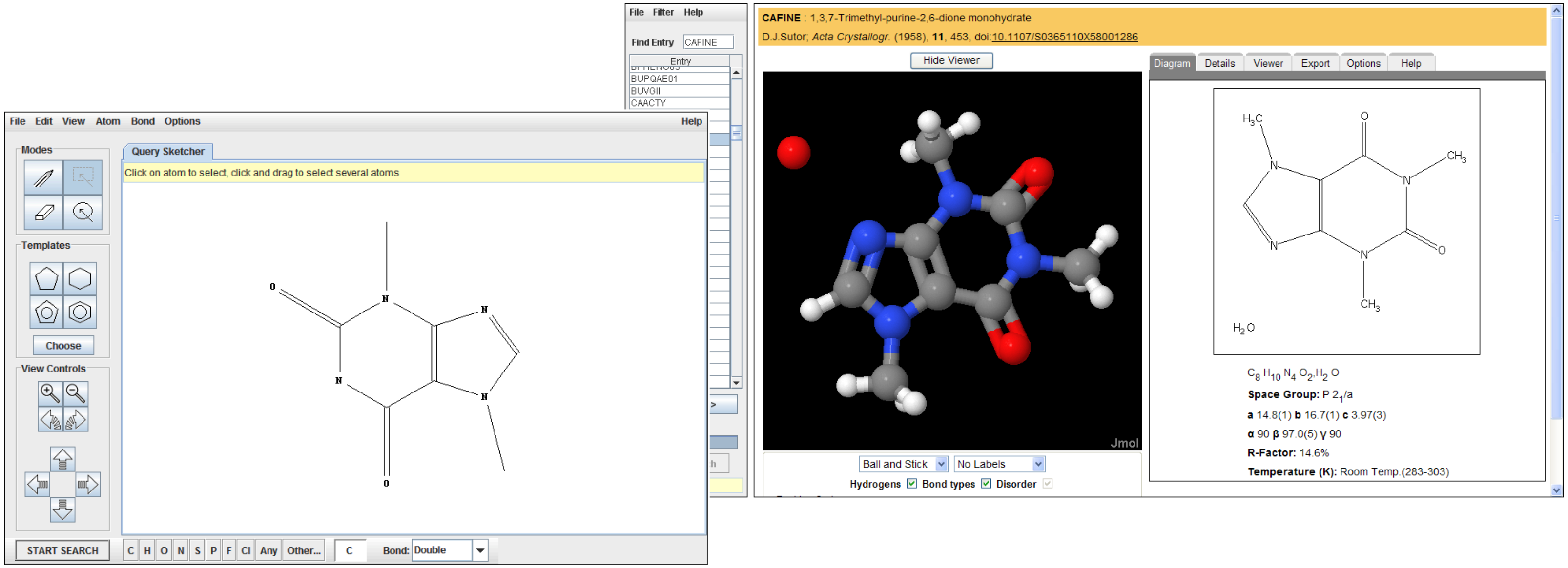
J. Chem. Educ., **2009**, 86 (4), p 460
DOI: 10.1021/ed086p460

*"**...**gives students a chance to move beyond their textbook structures and analyze data from actual molecular structures"*



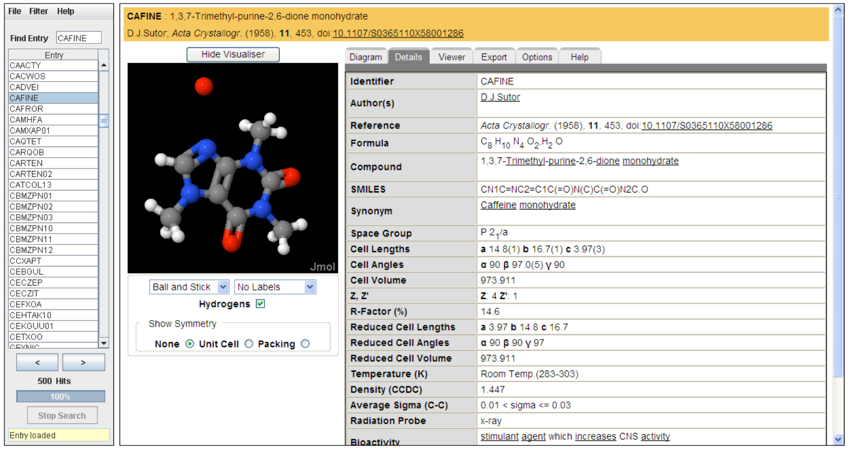
Improving CSD Accessibility

- **WebCSD**: online search interface to the CSD
- **University-wide access**: not required to download, install or register software on individual machines
- **Ease of use**: search and visualise structures using a standard web browser

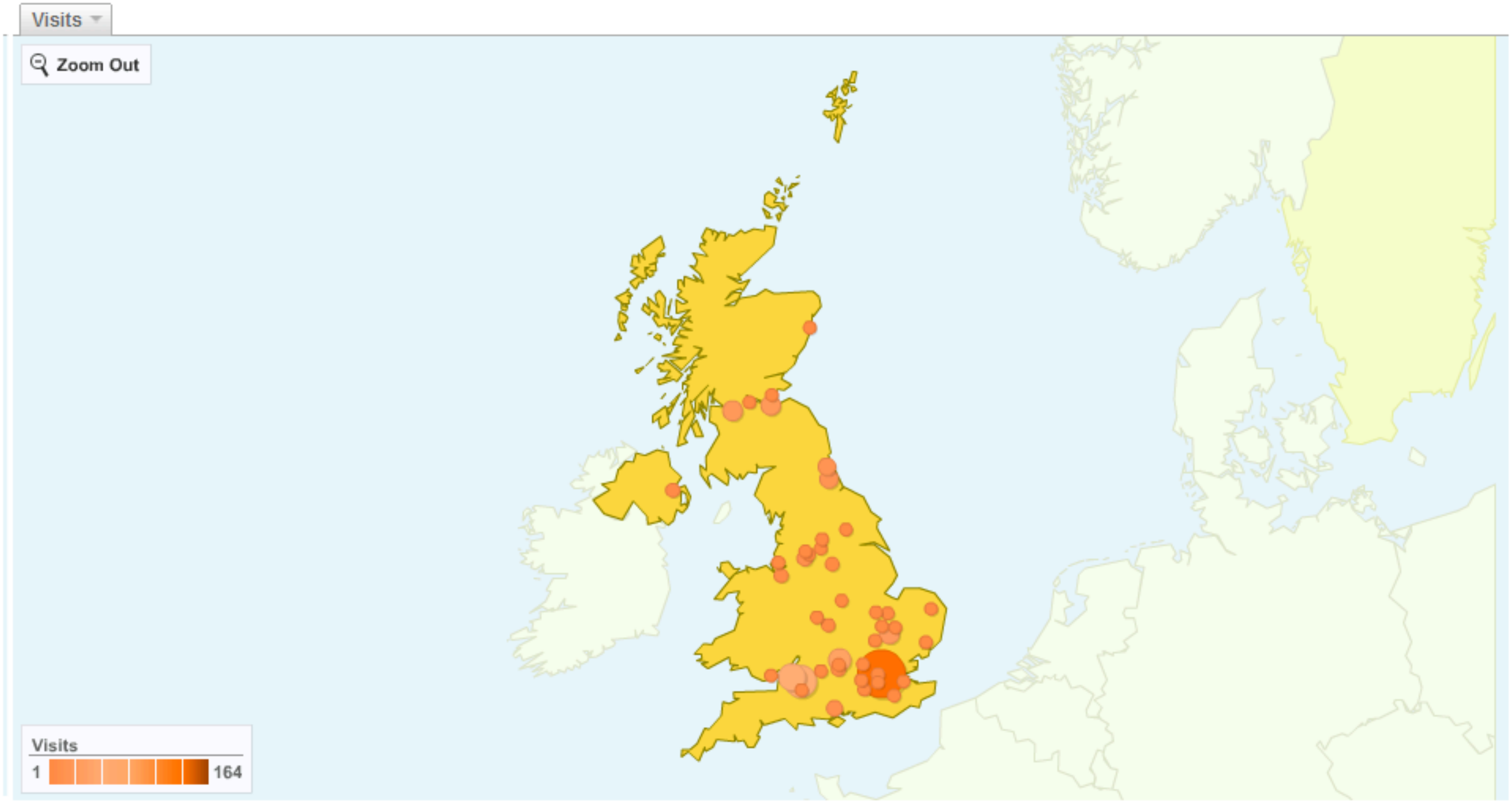


Search functionality

- Text and numeric
- Reduced cell
- 2D substructure
- 2D structure similarity
- View specific refcode or browse the CSD

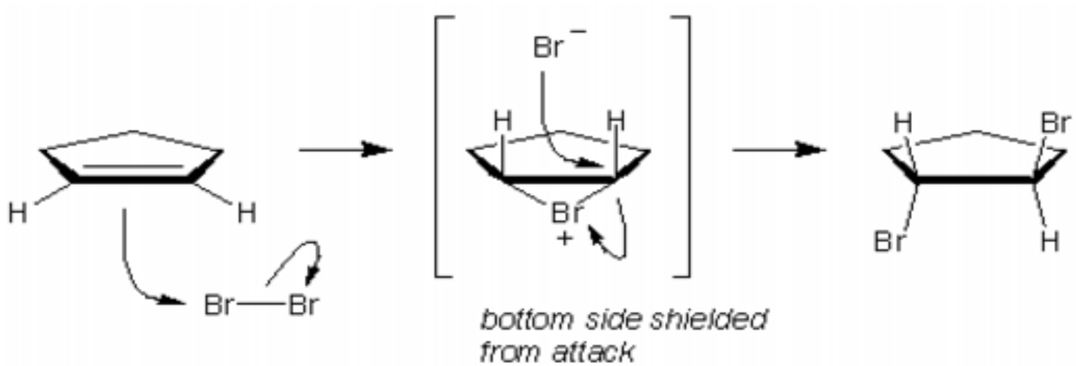


WebCSD Usage Jan - Mar 2011



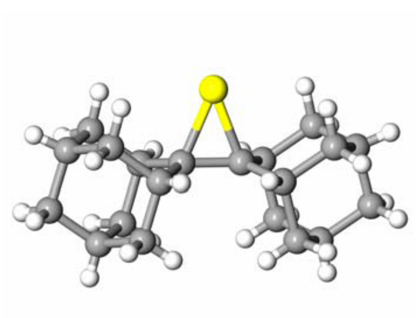
This country/territory sent 617 visits via 45 cities

Example 1: Exploring Reaction Mechanisms using Crystallographic Data

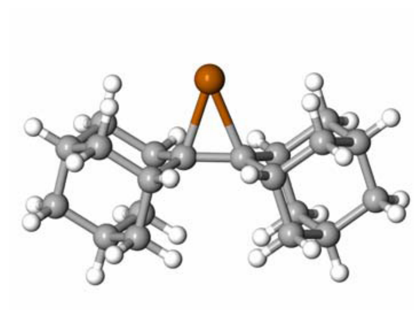


Learning Goals:

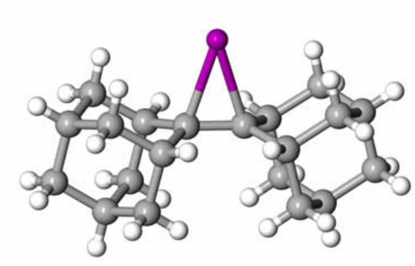
- Stereoselectivity
- Evidence for cyclic halonium ions
- Stabilisation



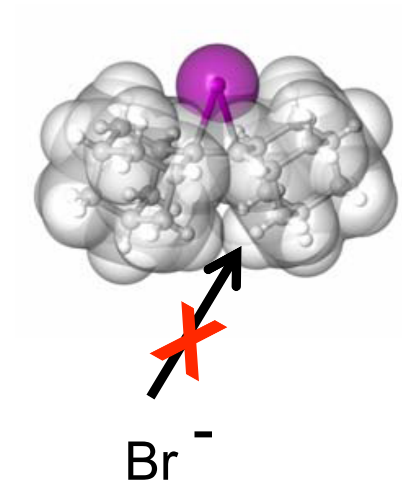
Cl - *UGIDP*



Br - *DAKVUG*



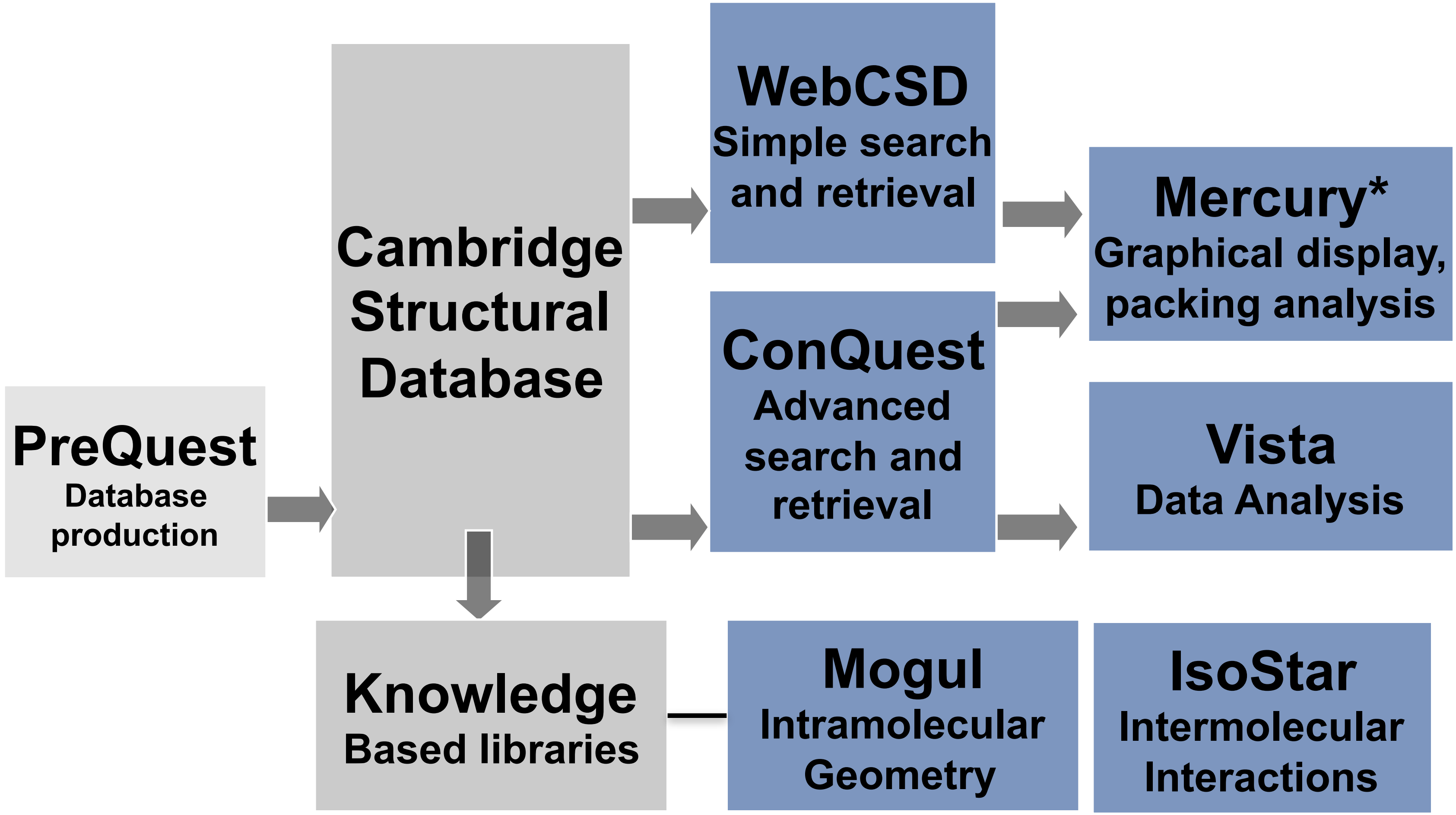
I - *WEVPOC*



http://www.ccdc.cam.ac.uk/free_services/teaching/

Database plus Access Software

Acta Cryst., B58, 380-388 & 389-397, 2002

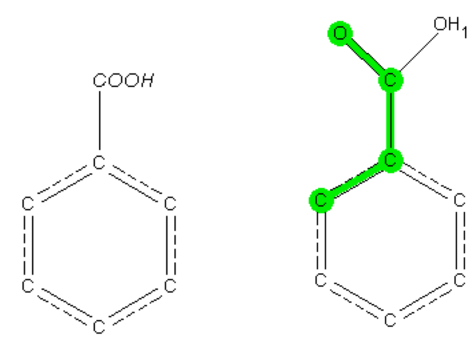


*free downloads http://www.ccdc.cam.ac.uk/

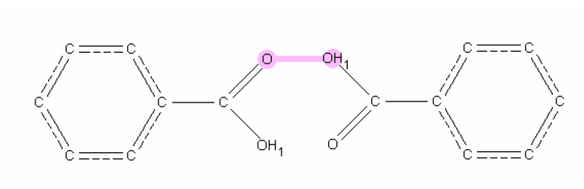
Searching the CSD: ConQuest

- Searches:
  - Text & numerical data

    Compound name, author name, Chemical formula, melting point, keyword…..
  - Chemical substructures in 2D/3D

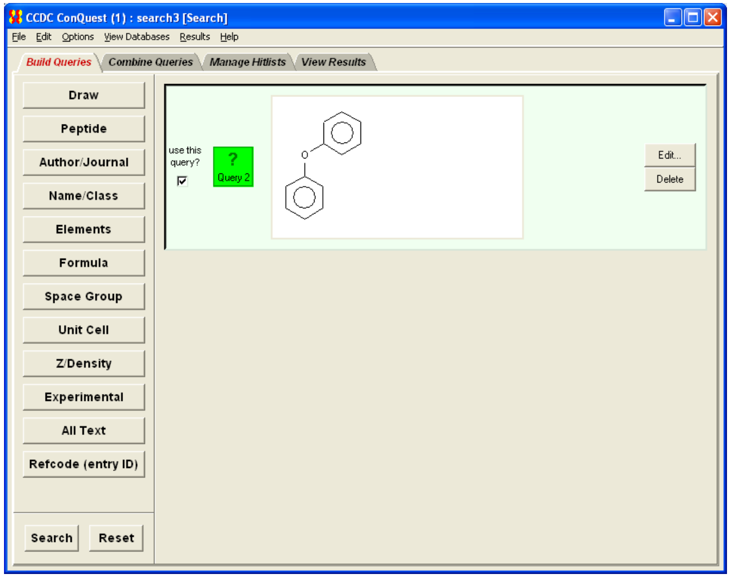


  - Intermolecular non-bonded contacts



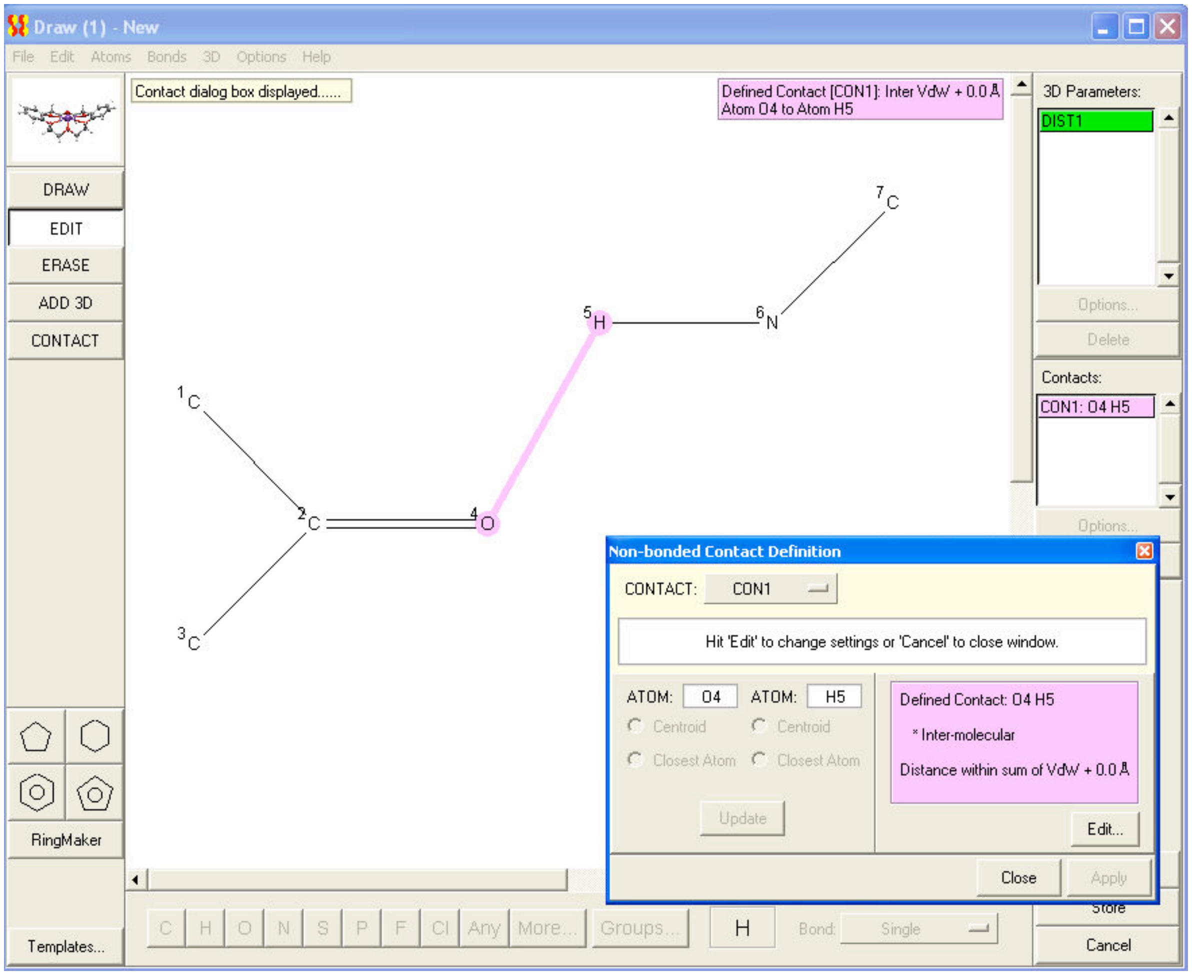
ConQuest

- Searches:
  - Access CSD and in-house data at the same time



  - Large range of filtering options based on experimental or chemical options
  - Combine searches and manage hitlists
  - Export results to Vista (numerical analysis), or Mercury (visualisation)

ConQuest Search:
N-H…O=C(amide) H-Bonds



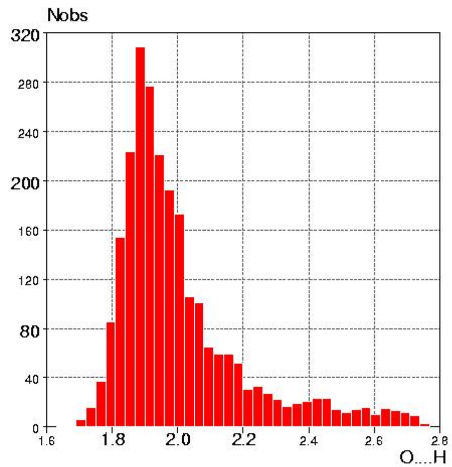
Search criteria:

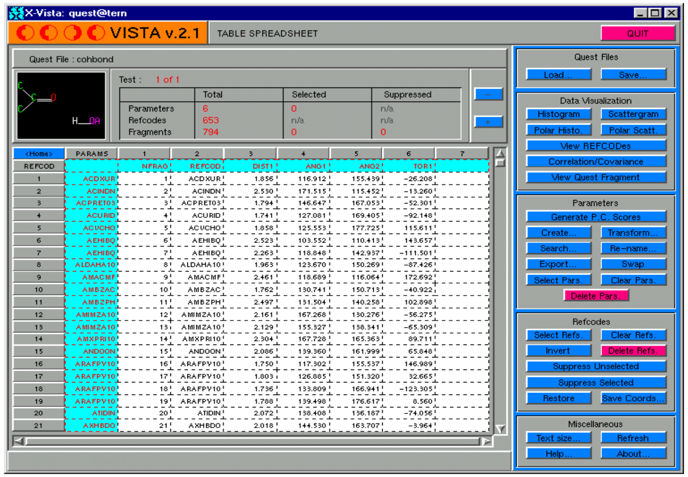
- C=O…H vdW+0.4
- Cryst. R-factor <5%

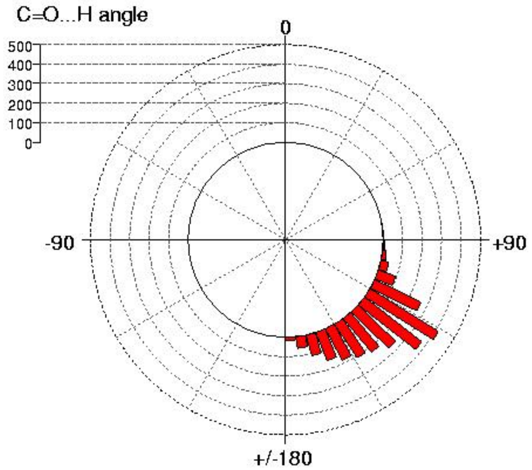
Calculate geometry:

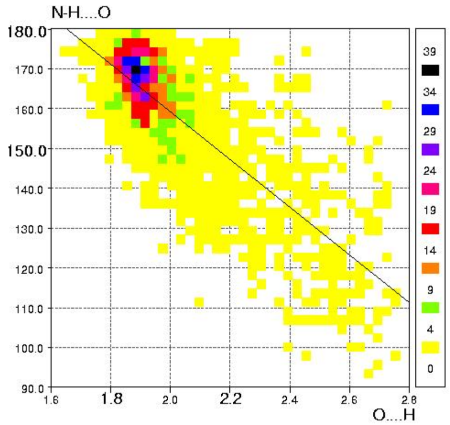
- O…H distance
- O…H-N angle
- C=O…H angle
- Angle between N-H vector and amide plane

Analysing Geometrical Data: Vista

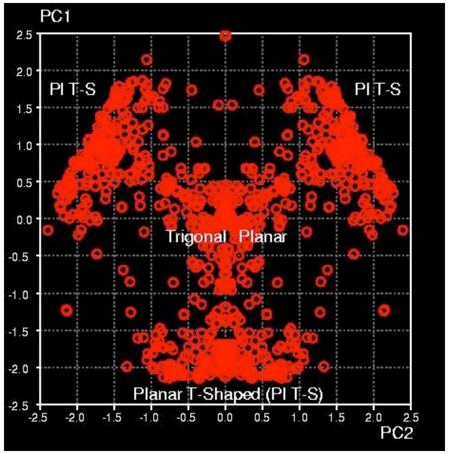




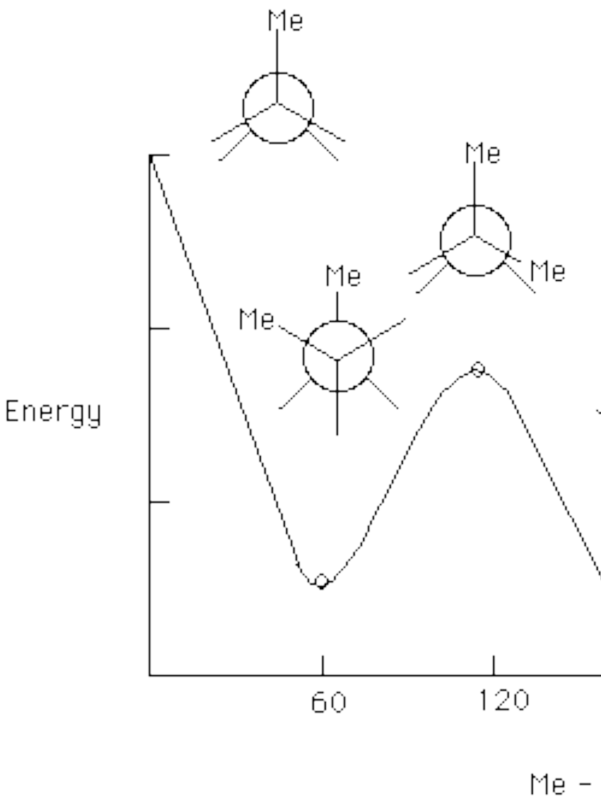


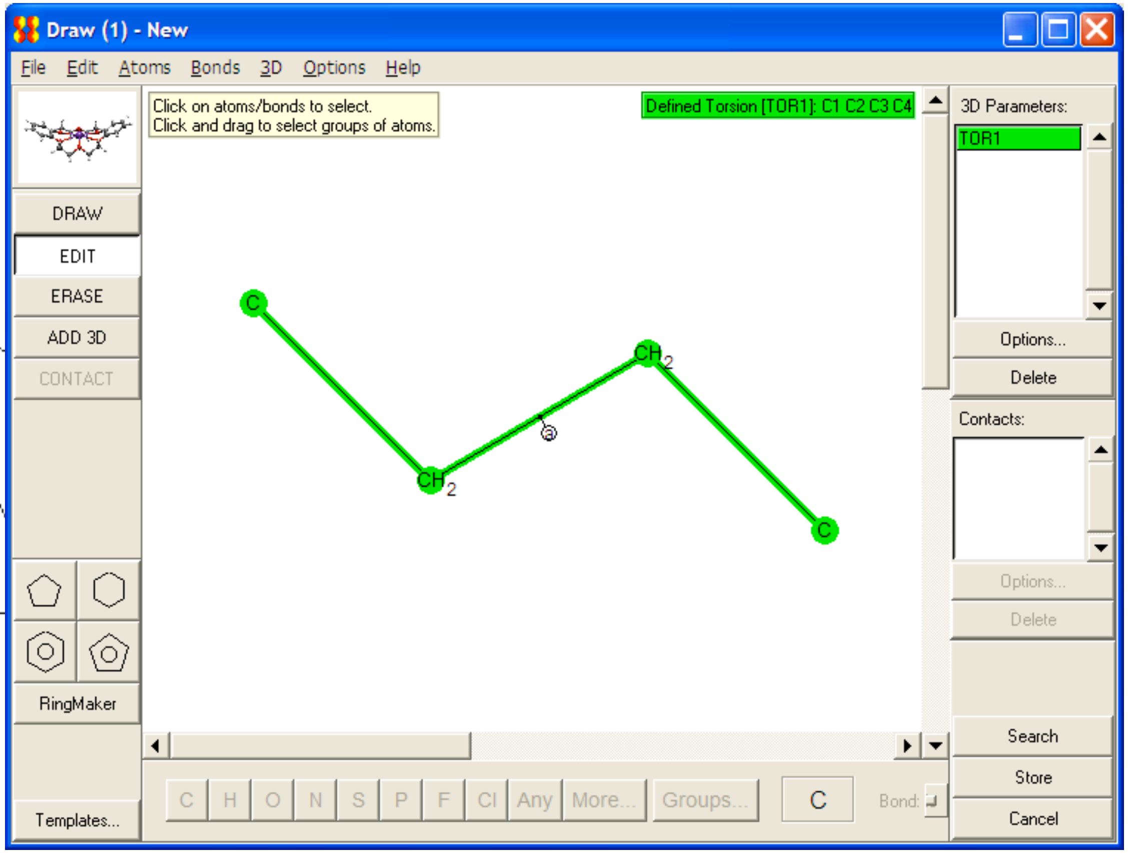


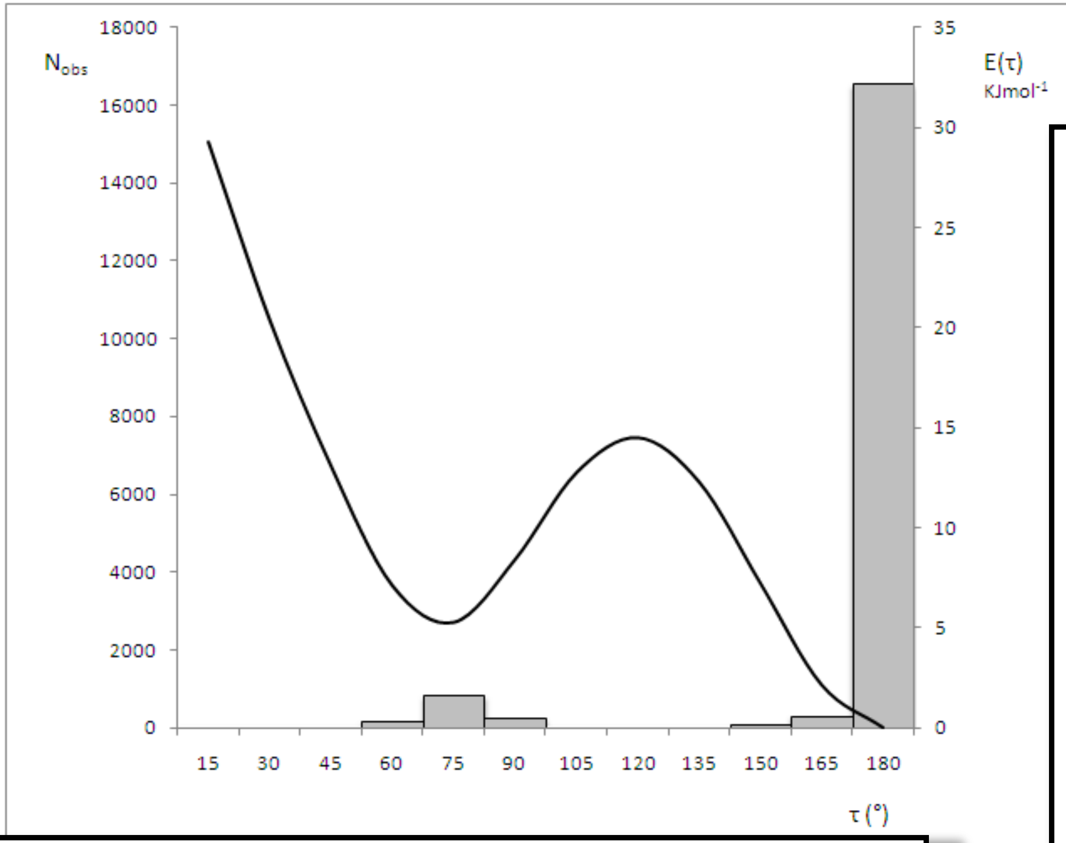
- *Histograms* (Cart^n, Polar)
- *Scatterplots* (Cart^n, Polar)
- Apply *mathematical functions* to spreadsheet parameters
- *Statistical analysis*:
  - Means, medians,esd's etc. for parameter distributions
  - Regression
  - Principal components (PCA)

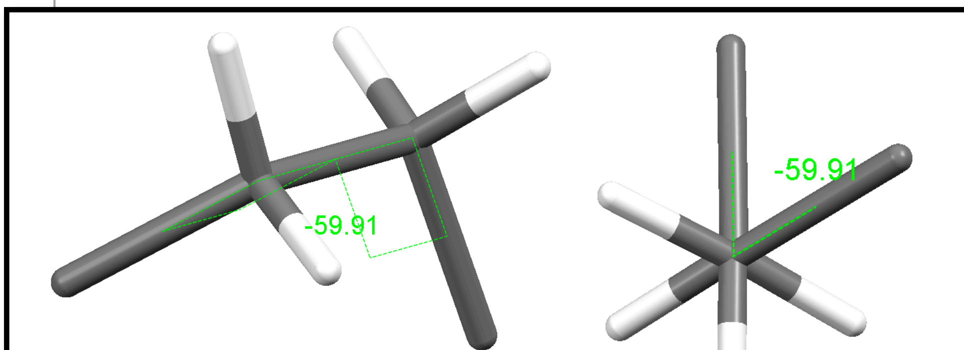


Example 2: Conformational Analysis

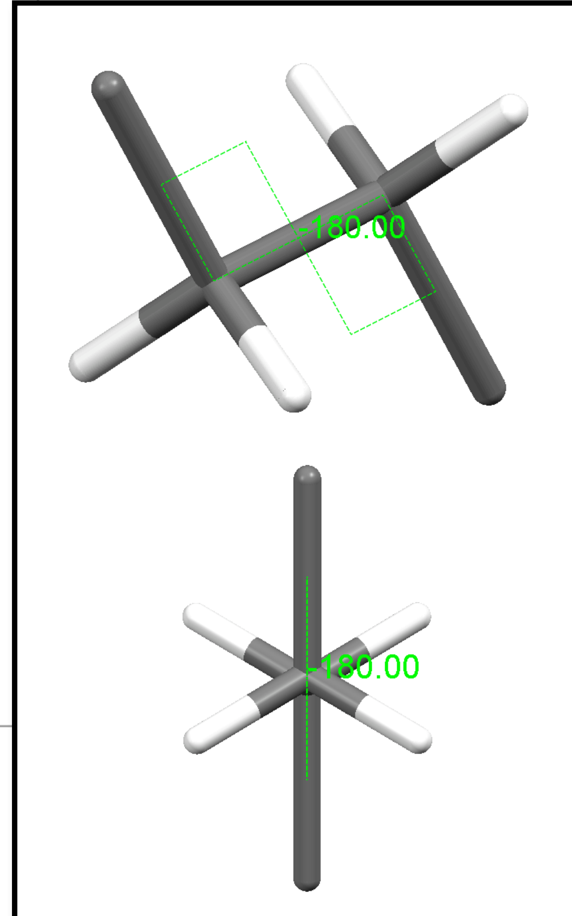








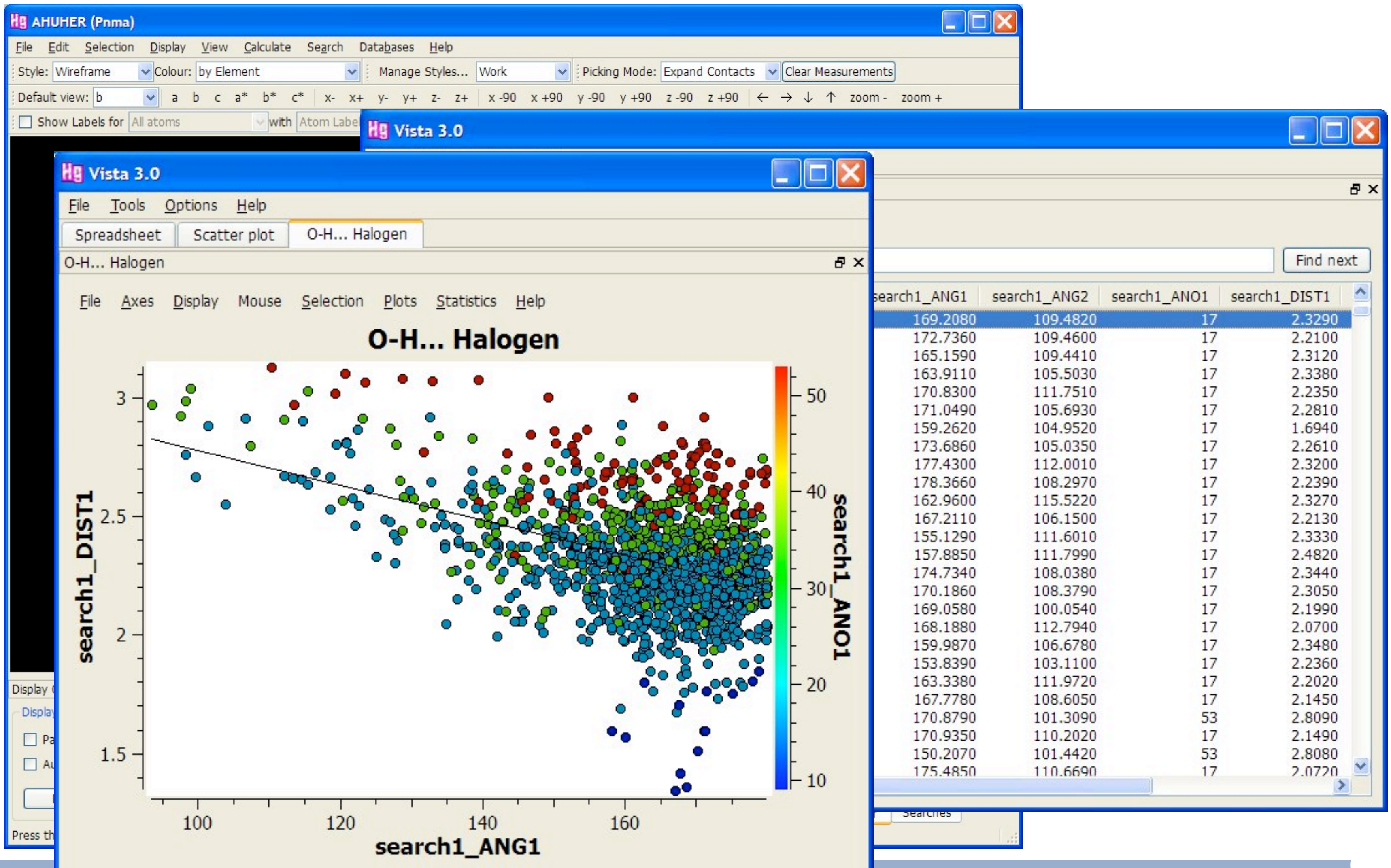
Gauche (ABIPIK)



Anti (DUCKOB04)

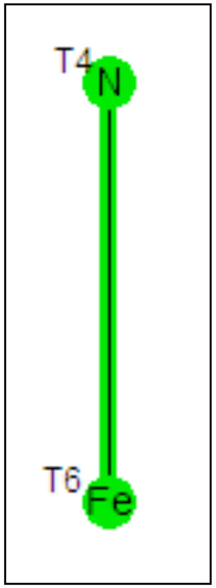
Replacing Vista

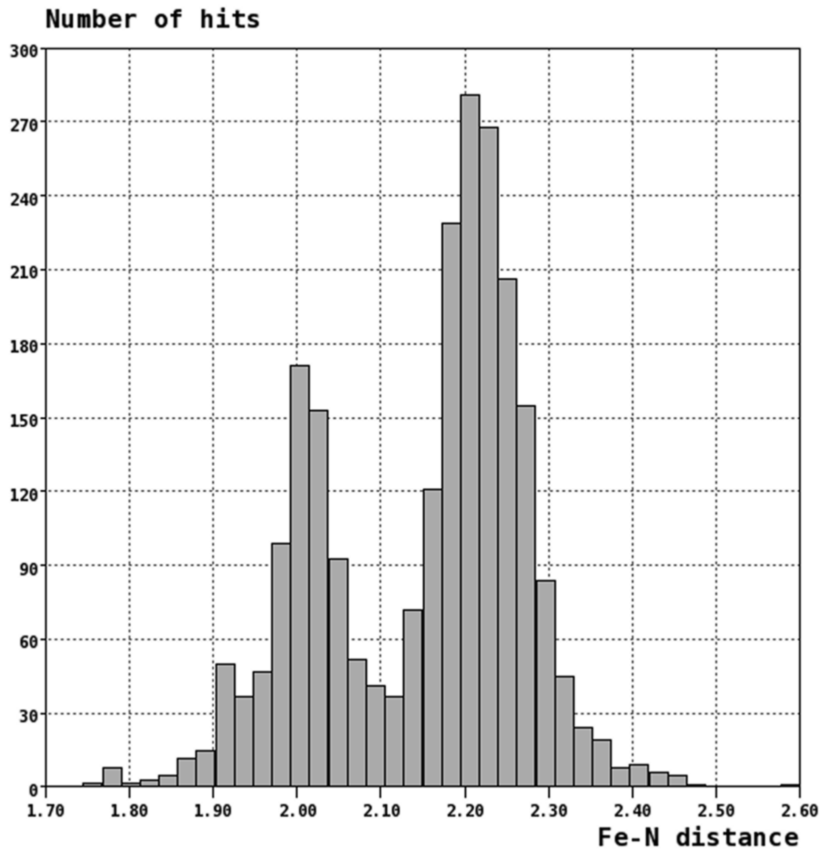
Interactive display and analysis of numeric CSD search data



earch1_ANG1	search1_ANG2	search1_ANO1	search1_DIST1
169.2080	109.4820	17	2.3290
172.7360	109.4600	17	2.2100
165.1590	109.4410	17	2.3120
163.9110	105.5030	17	2.3380
170.8300	111.7510	17	2.2350
171.0490	105.6930	17	2.2810
159.2620	104.9520	17	1.6940
173.6860	105.0350	17	2.2610
177.4300	112.0010	17	2.3200
178.3660	108.2970	17	2.2390
162.9600	115.5220	17	2.3270
167.2110	106.1500	17	2.2130
155.1290	111.6010	17	2.3330
157.8850	111.7990	17	2.4820
174.7340	108.0380	17	2.3440
170.1860	108.3790	17	2.3050
169.0580	100.0540	17	2.1990
168.1880	112.7940	17	2.0700
159.9870	106.6780	17	2.3480
153.8390	103.1100	17	2.2360
163.3380	111.9720	17	2.2020
167.7780	108.6050	17	2.1450
170.8790	101.3090	53	2.8090
170.9350	110.2020	17	2.1490
150.2070	101.4420	53	2.8080
175.4850	110.6690	17	2.0720

Example 3:
High spin vs. low spin complexes

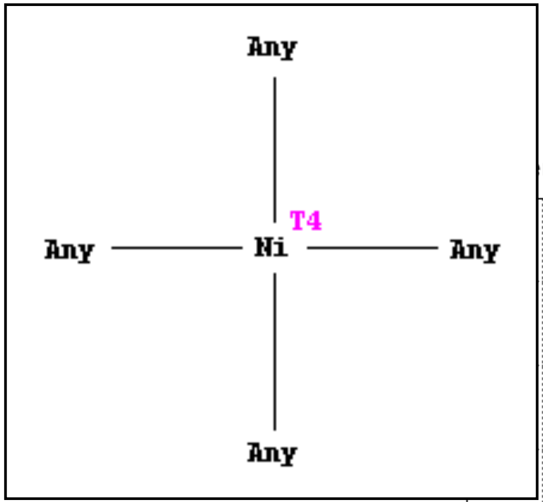


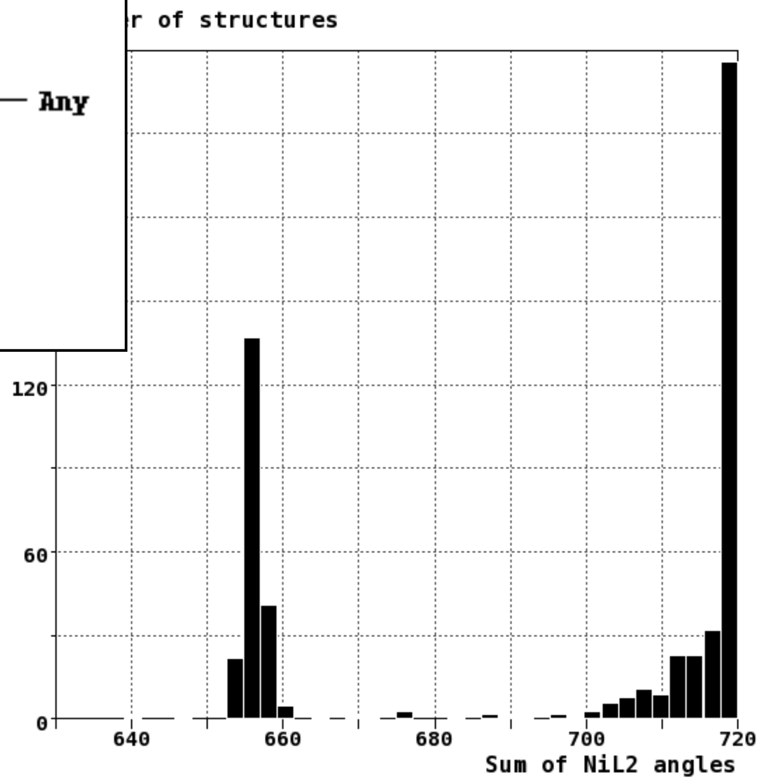


High spin and low spin forms of octahedral Fe(III)

cf. octahedral Ni(II)

Example 3:
Ni coordination geometries





Sum of L-Ni-L angles showing square planar and tetrahedral geometries

Further Information

- Teaching 3D Structural Chemistry Using Crystal Structure Databases 1. An Interactive Web Accessible teaching Subset of the Cambridge Structural Database
  Gary M. Battle and Frank H. Allen, Gregory M. Ferrence
  ***J. Chem. Educ.*, 2010, 87 (8), pp 809–812**

- Teaching 3D Structural Chemistry Using Crystal Structure Databases 2. Teaching Units that Utilize an Interactive Web Accessible teaching Subset of the Cambridge Structural Database
  Gary M. Battle and Frank H. Allen, Gregory M. Ferrence
  ***J. Chem. Educ.*, 2010, 87 (8), pp 813–818**

- Applications of the Cambridge Structural Database in Chemical Education
  Gary M. Battle and Frank H. Allen, Gregory M. Ferrence
  ***J. Appl. Cryst. Special Teaching Edition*, 2010, 43, 1208-1223**

Free Educational Resources

- Teaching subset of 500 CSD entries chosen to illustrate a wide range of 3D structural issues
- Web-based interface for browsing the teaching subset
- Tutorials, class exercises and worksheets

http://www.ccdc.cam.ac.uk/free_services/teaching/

NSF Discovery Corps Fellowship Grant No. 0725294

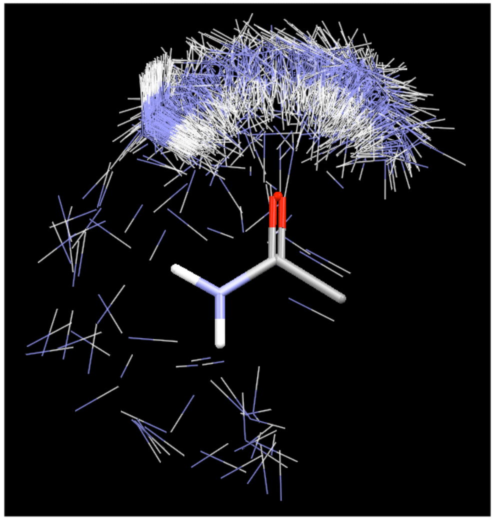
Part #2

- IsoStar and SuperStar
  - IsoStar: Assessing non-bonded interactions
  - SuperStar: using IsoStar data to predicting binding site interactions

IsoStar

A Knowledge Base of Intermolecular Interactions

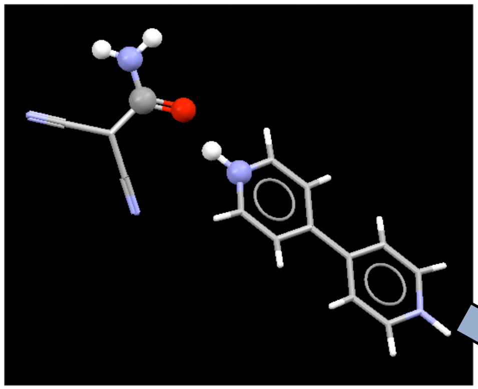
- Experimental data taken from:
  - Cambridge Structural Database
  - Protein Data Bank (protein-ligand complexes only)
- Interaction distributions displayed as scatterplots or contour surfaces

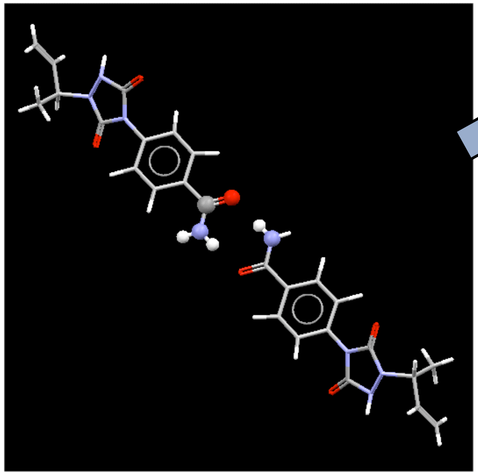


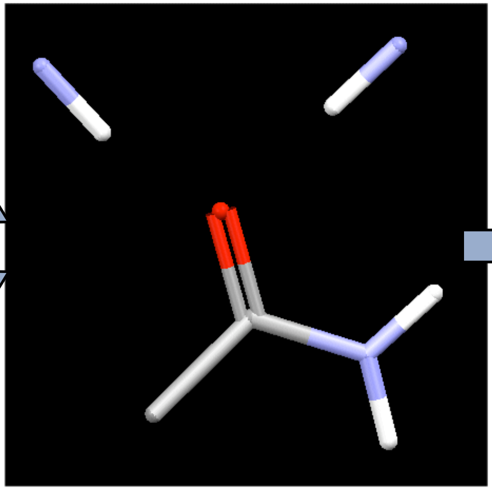
IsoStar Methodology

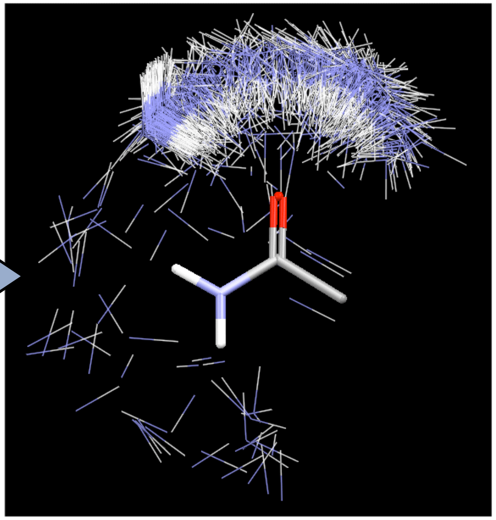
central group: -$CONH_2$

contact group: NH





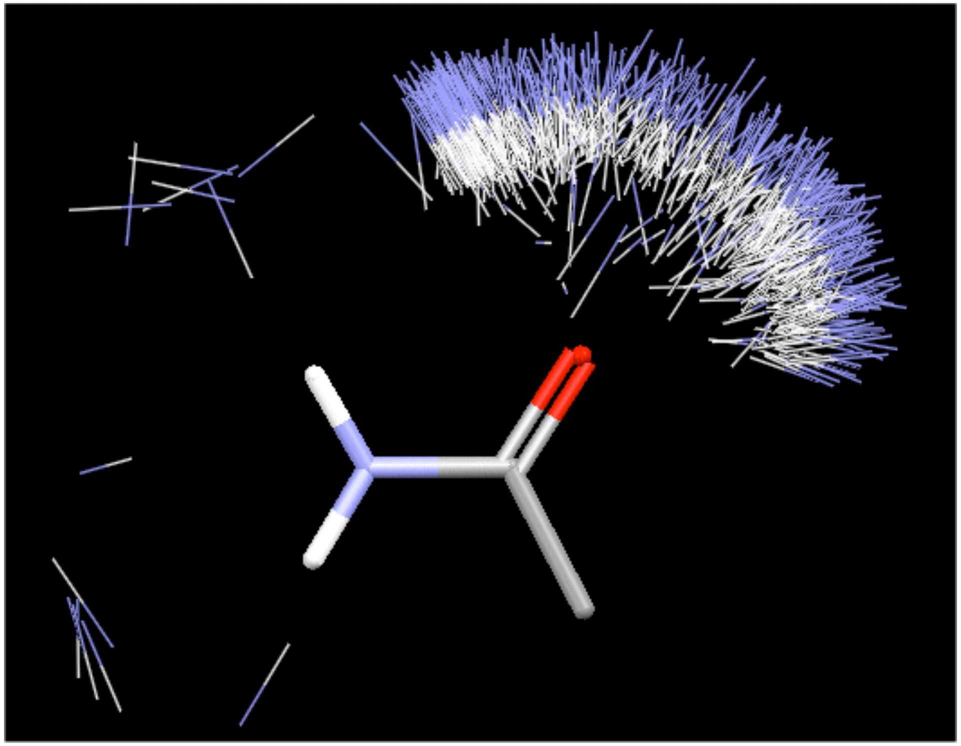


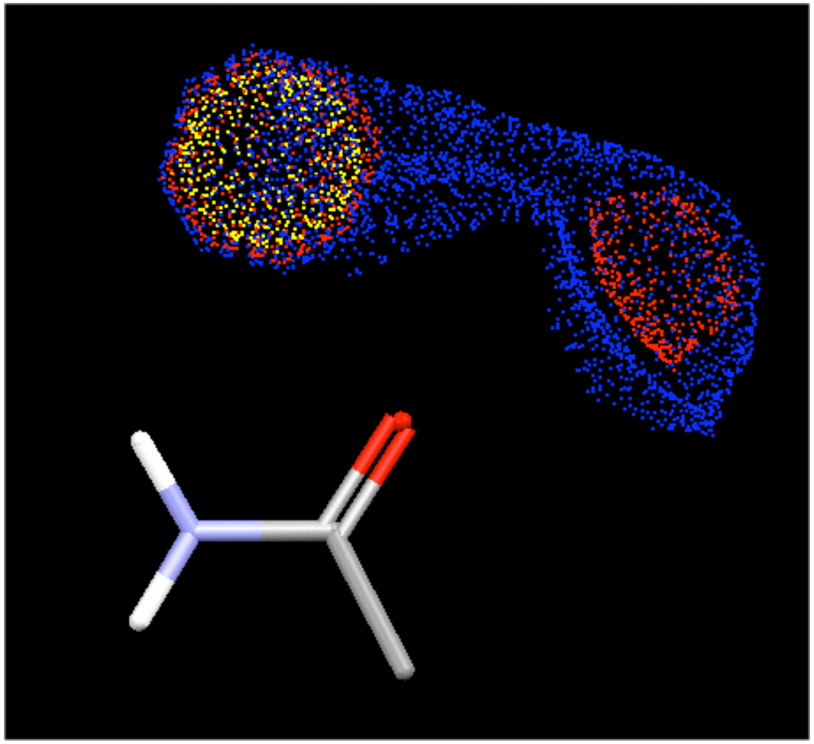


Search CSD or PDB for structures containing desired contact

Superimpose hits and display as scatterplots

Density Maps





Can also represent distribution as density maps

Coverage of Groups

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

- Experimental data from CSD and PDB protein-ligand complexes
- >300 central groups, >50 contact groups
- >22,000 CSD scatterplots
- >7,400 PDB scatterplots
- 1,550 theoretical energy minima from DMA/IMPT

Generate your own scatterplots using IsoGen

Typical Uses of IsoStar

Probability of an interaction occurring

Will fluorine act as an H-bond acceptor?

Is the N or the O of oxazole more likely to H-bond?

Preferred geometries

Do H-bonds to ether O lie along lone pairs?

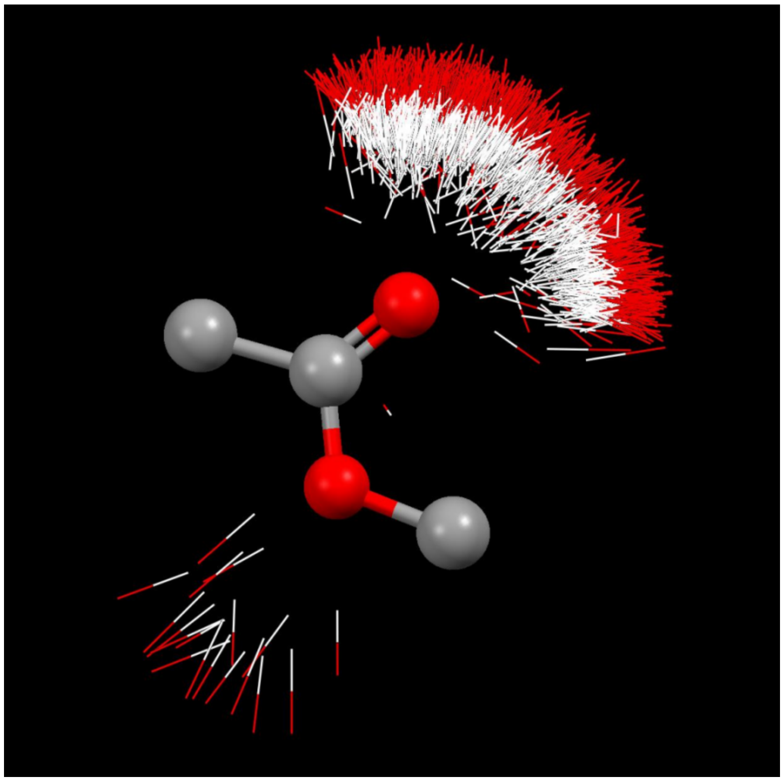
Do aromatic rings stack or form T-shape arrangements?

Design strategies

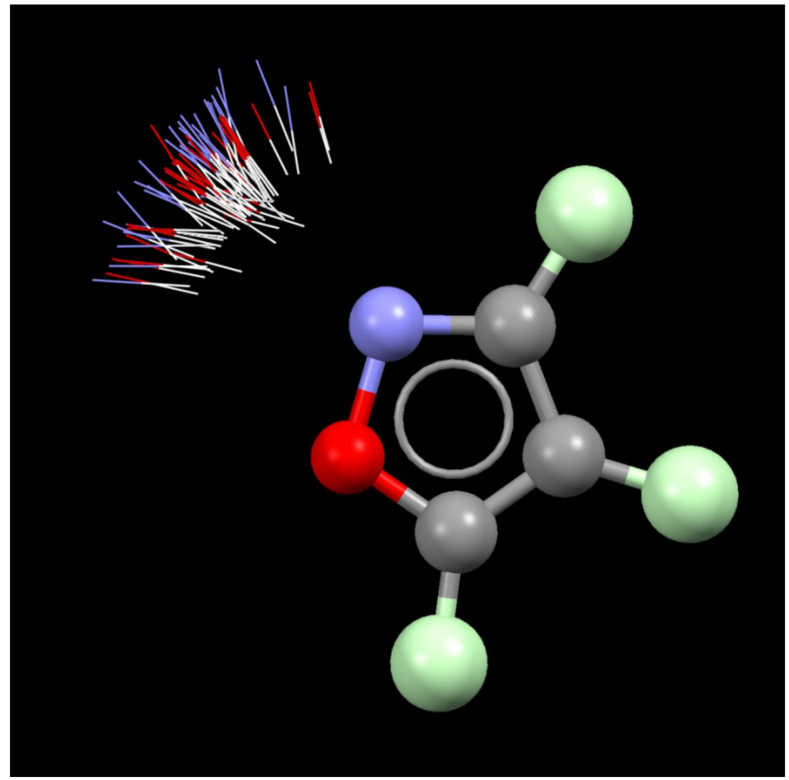
Do thiazole and oxazole form similar interactions?

How can we bind to a Trypsin residue?

Strong acceptors compete for donor-H groups

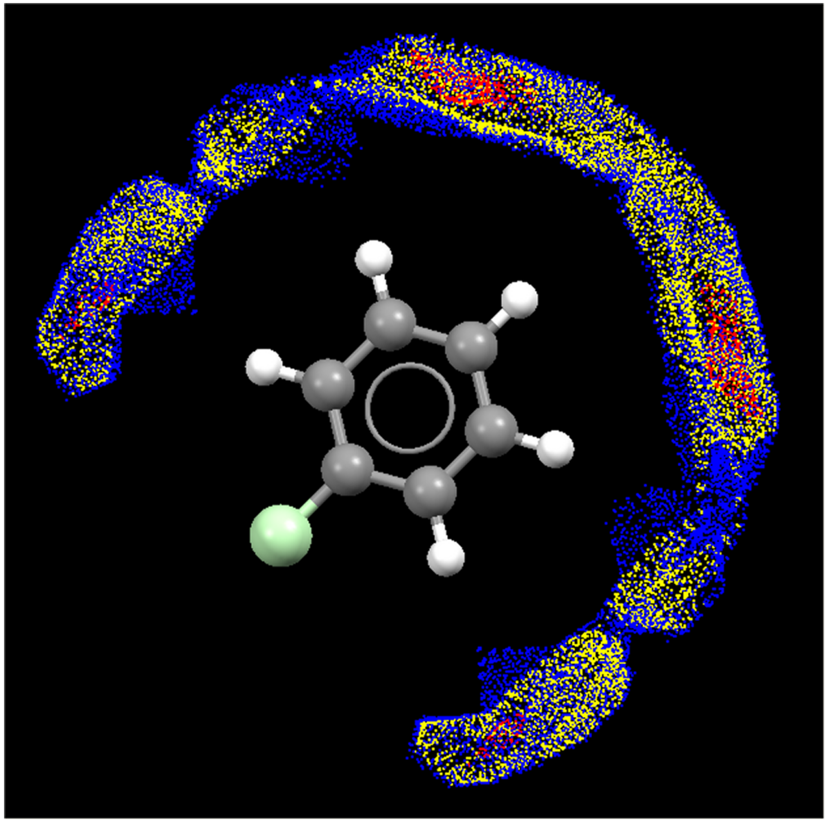


Scatterplot of O-H…O H-bonds in esters

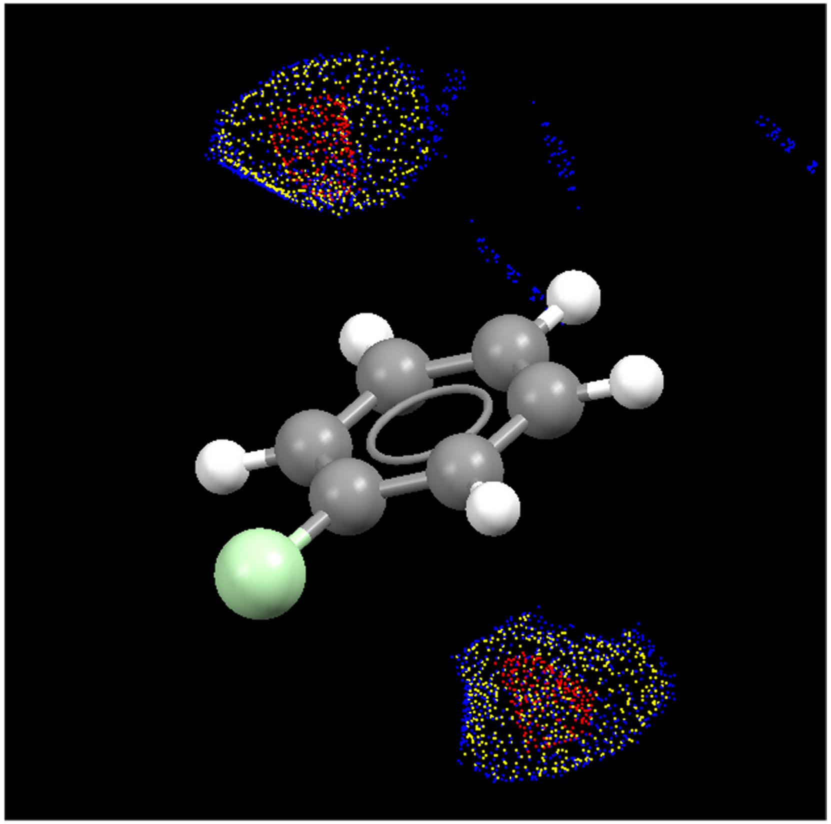


O-H and N-H H-bonds to N- and O-acceptors in isoxazole rings

Electronegativity and electropositivity



carbonyl (–C=O$^{\delta-}$) groups around a phenyl (–C_6H_5) central group



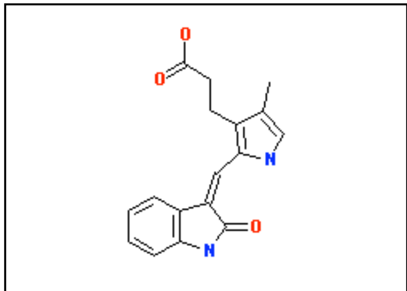
alkyl C-H around a phenyl (–C_6H_5) central group

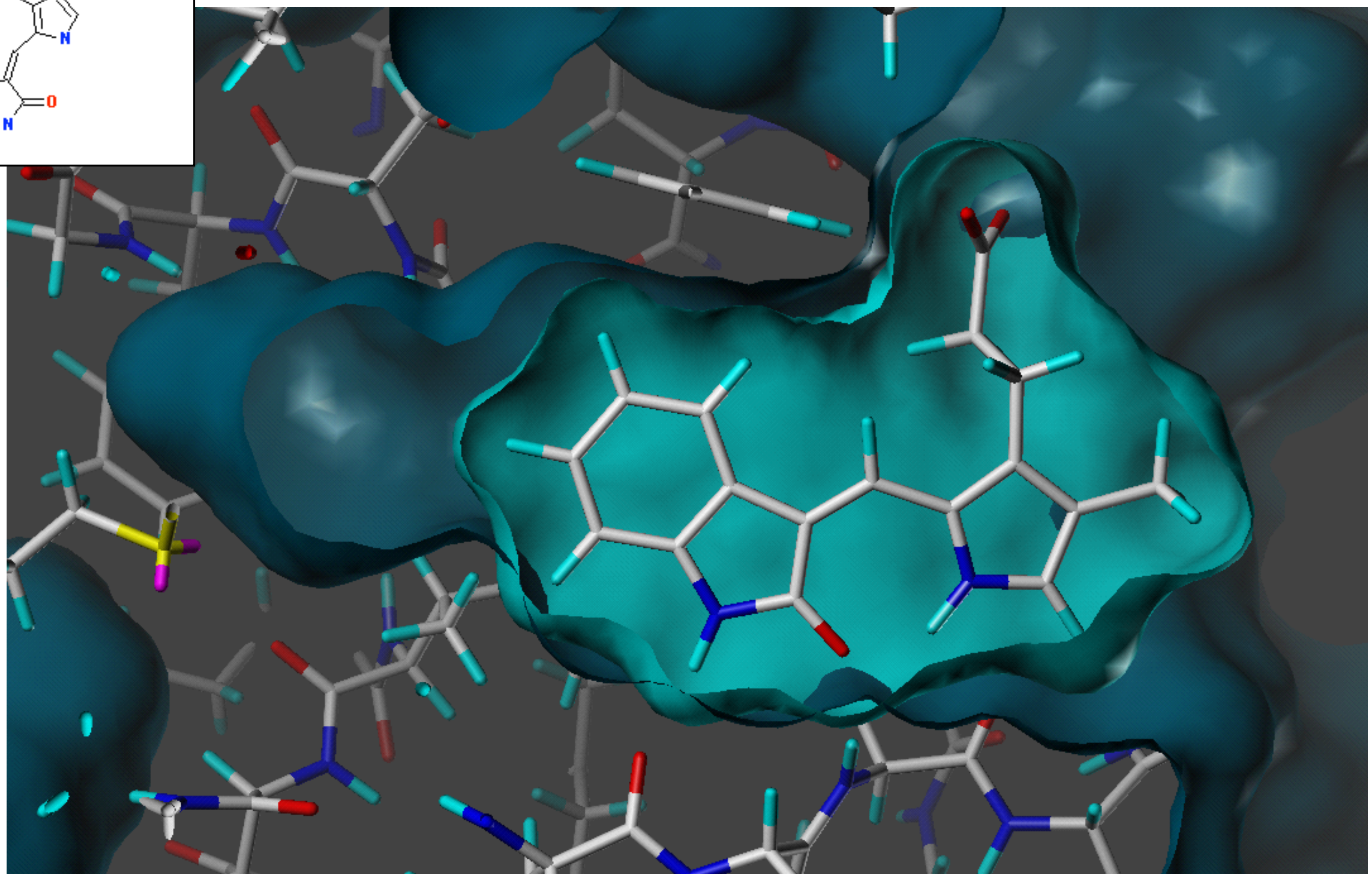
SuperStar

Knowledge-base can provide input for other software

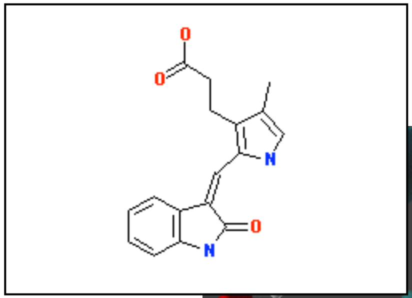
- SuperStar is a program for identifying interaction sites in proteins or around small molecules
- Predicts where a given functional group ('probe') is likely to interact at binding site
- Based exclusively on experimental data in IsoStar
- Validated successfully on >300 protein-ligand complexes from PDB
- Uses include pharmacophore generation

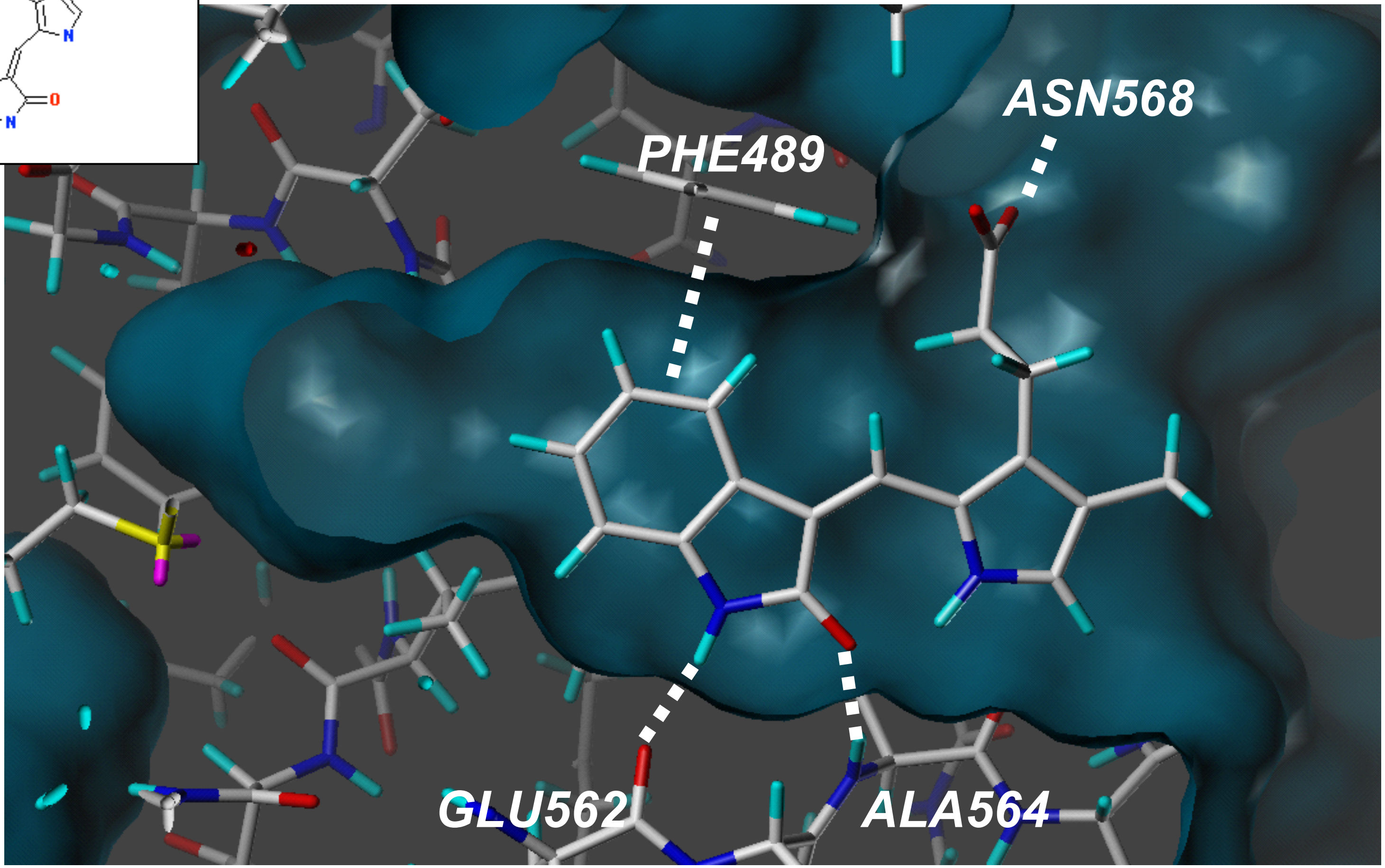
Tyrosine kinase (*1fgi*)



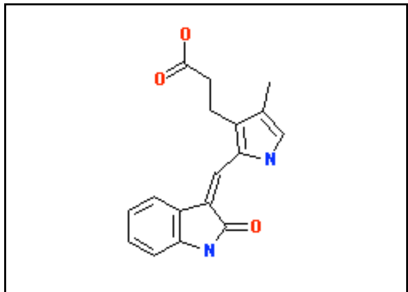


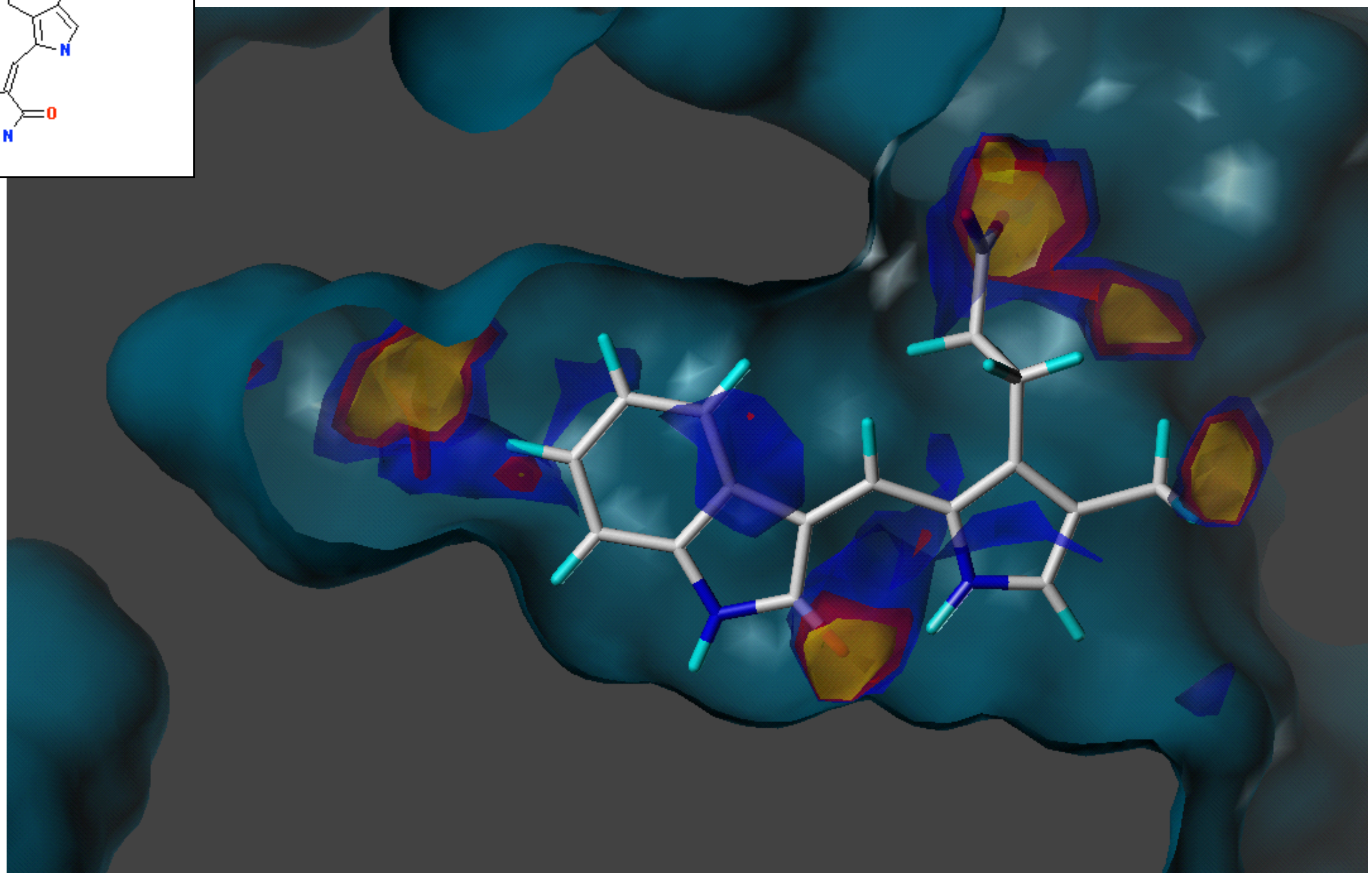
Tyrosine kinase (*1fgi*)



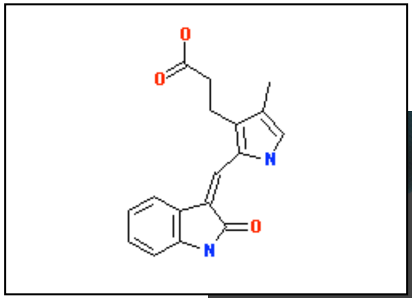


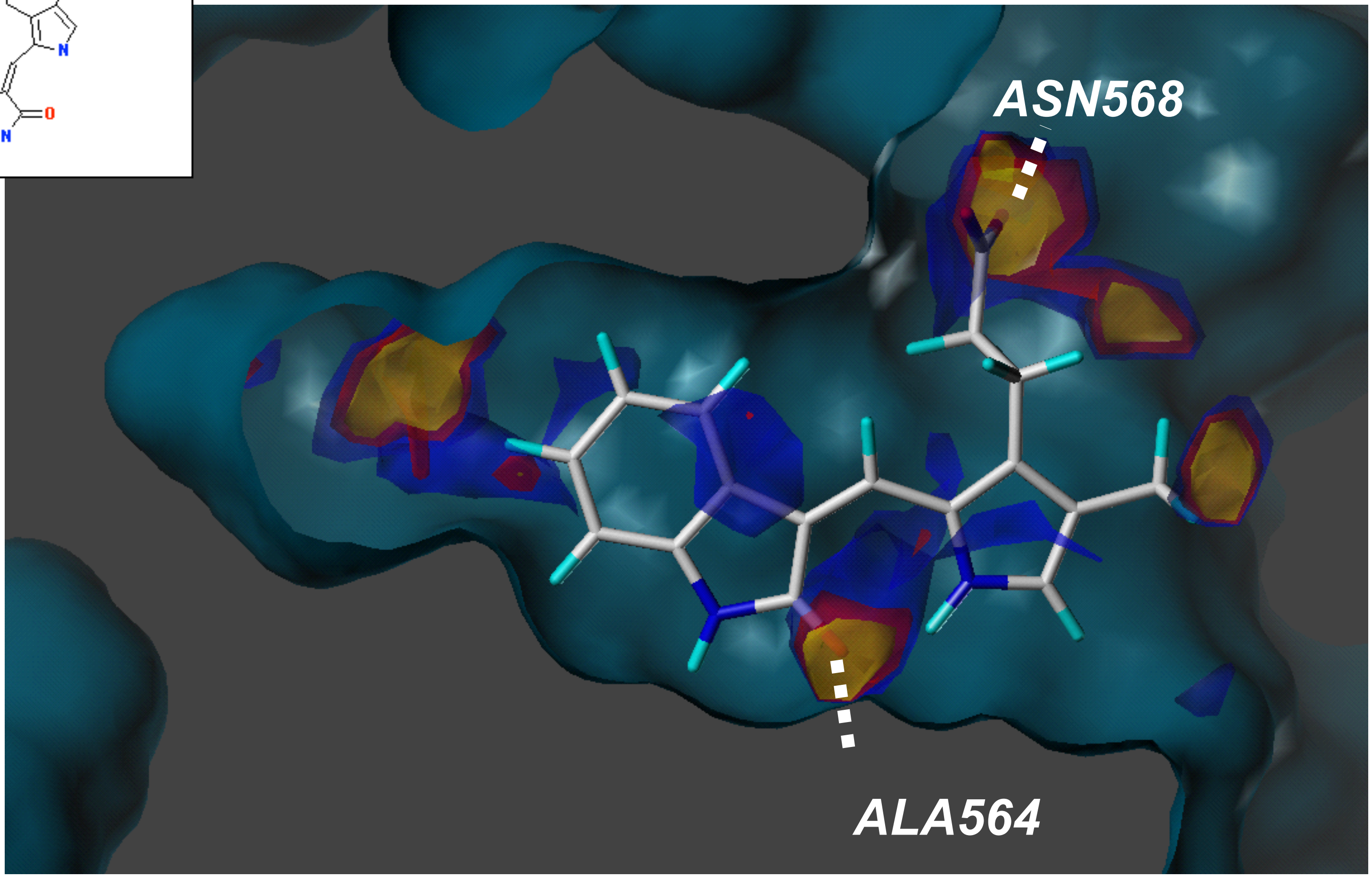
Map for CO Oxygen Probe



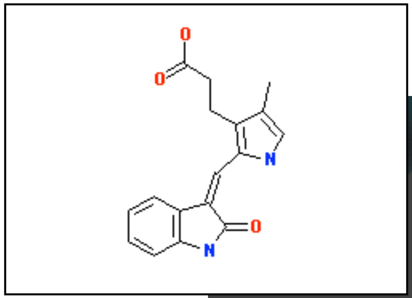


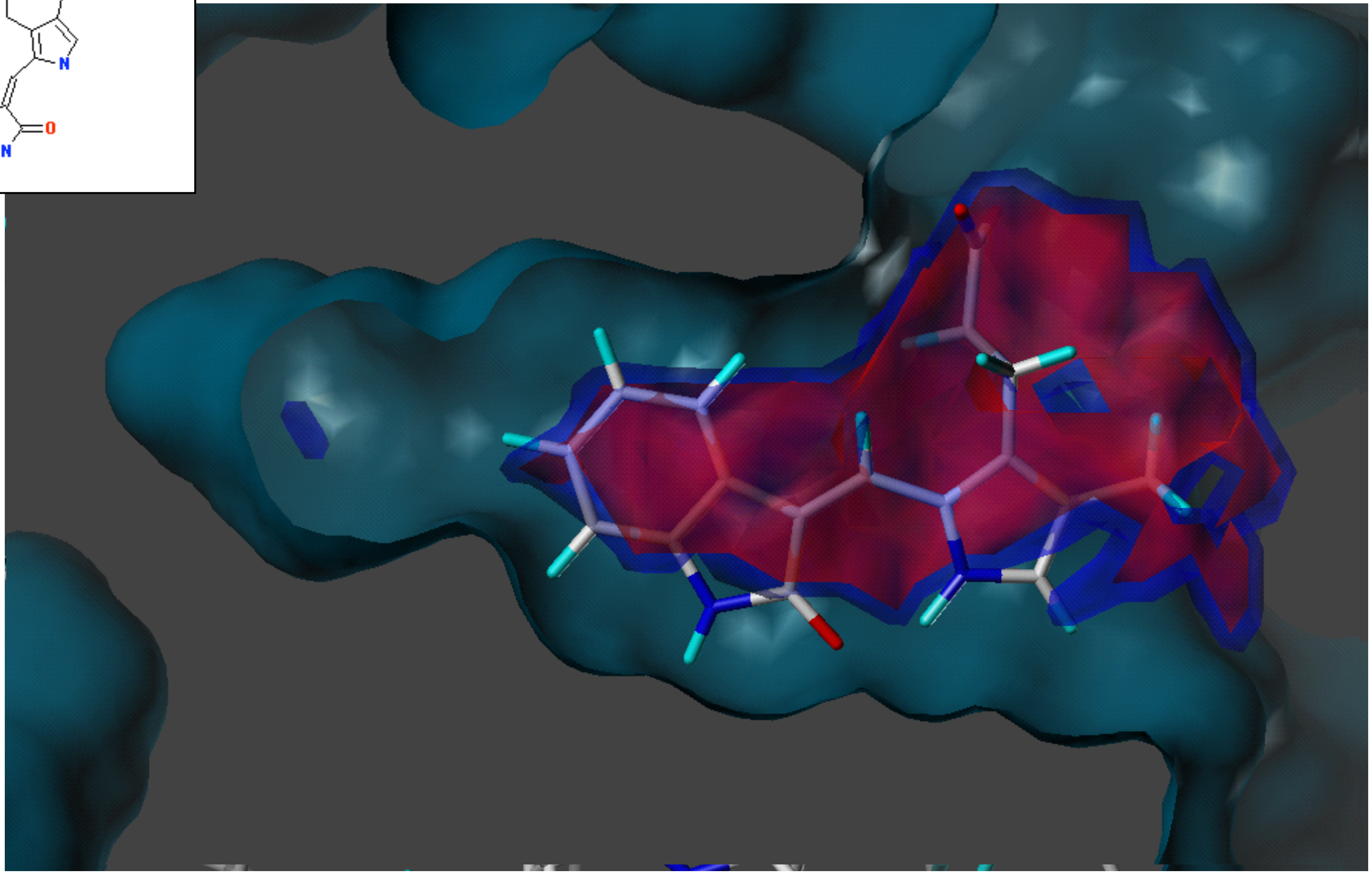
Map for CO Oxygen Probe





Map for aromatic CH carbon probe





Thank you!

- Email us…
  [*Teaching@ccdc.cam.ac.uk*](mailto:Teaching@ccdc.cam.ac.uk)
- Follow us…
  www.facebook.com/ccdc.cambridge

Acknowledgement -

NSF Discovery Corps Fellowship Grant No. 0725294