# Why Validation has become an Issue

## **DIY**

Do-it-yourself crystallography is now a reality. Instruments like the Bruker-Nonius kCCD and the Collect-Denzo-Scalepack suite of programs make reliable X-ray data collection quite practical with very little training.

# Why Validation has become an Issue

Structure solution and refinement programs have also become less difficult to use.

In fact, SIR2002 and SHELXS are now so refined, there is probably very little the casual user can do if they fail to solve a structure.

The chemist now has a real possibility of determining their own crystal structures.

# Why Do-it-Yourself Analyses?

**Economics**

We estimate that, excluding floor charges, depreciation etc, the actual running cost of a kCCD diffractometer is 12,000 Euro p.a.

This covers:

Electricity, Liquid Nitrogen, X-ray tube, routine maintenance, air conditioning, chilled water.

# Why Do-it-Yourself Analyses?

**Economics**

In Oxford, except for special studies, the machines are run on a 3-shift system:

| | |
|---|---|
| 9.00-13.00 | 4 hours |
| 13.00-1700 | 4 hours |
| Overnight | 16 hours |

This gives a capacity of over 900 crystal examinations pa, i.e. 13 Euro per experiment.

# Why Do-it-Yourself Analyses?

**Economics**

In the UK, a young research assistant earns about 35,000 Euro pa.

Given a mix of easy and difficult organic & organo-metallic structures, a sustainable throughput might be 200 experiments pa, making the man-power cost per structure 175 Euro.

# Why Do-it-Yourself Analyses?

**Economics**

Thus we have an instrument with a capacity of 900 investigations pa, and a research assistant with a capacity of 200 investigations pa. What is to be done with the remaining time?

A common solution is to reduce the throughput to only one investigation each day.

# DIY Crystallography – Good or Evil?

## The Good

- Improve chemists awareness of power and limitations of X-ray analysis.
- Make best use of crystallographers time.
- Make efficient use of diffractometers.

## The Bad

- Inexperienced analysts can make mistakes.
- Careless users could wreck machine.

# Chemical Validation

*The aim of most structure analyses is to investigate physical or chemical problems. At the end of an analysis the questions to be asked are:*

Is there any significant possibility that the structure may be ‘wrong’?

What confidence can we have in the molecular parameters?

What confidence can we have in the adps?

# Chemical Validation

If a structure ‘looks wrong’, it probably ***is*** wrong.

**‘X-ray analysis is still an art, through which our way is lighted by crystal-chemical knowledge, and chemical intuition.’**

The Cambridge Crystallographic Data Base stores much of our knowledge of what ‘looks right’, and MOGUL is a tool for accessing it.

Richard Cooper
www.xtl.ox.ac.uk

# What is MOGUL?

A data base of molecular geometries.

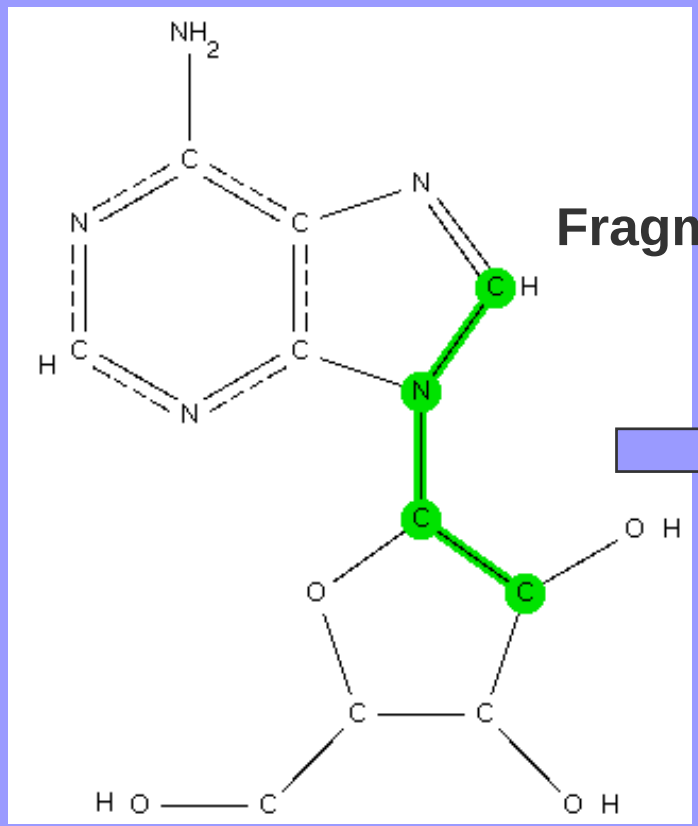
For every distinguishable chemical fragment in the CDS, the mean, standard deviation and distribution is computed.

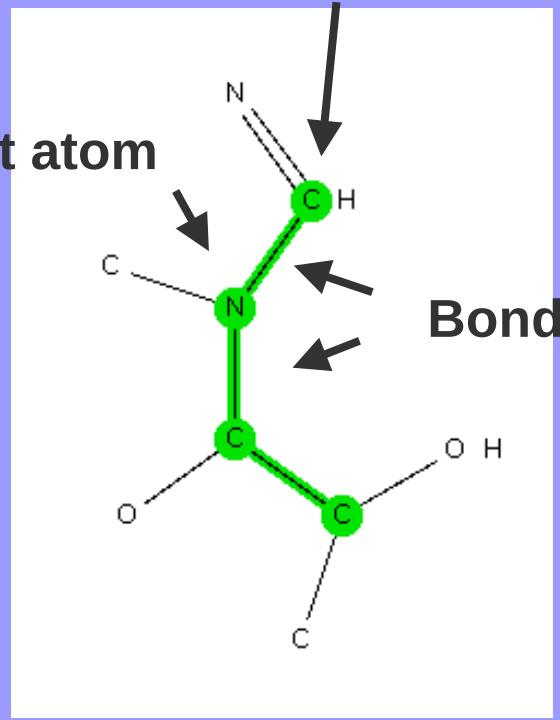
Because the CSD has been pre-processed, MOGUL has super fast search speeds (a few seconds) by avoiding graph matching.

Richard Cooper
www.xtl.ox.ac.uk

# How does it do that?

By building a library of chemical fragments.



Richard Cooper
www.xtl.ox.ac.uk

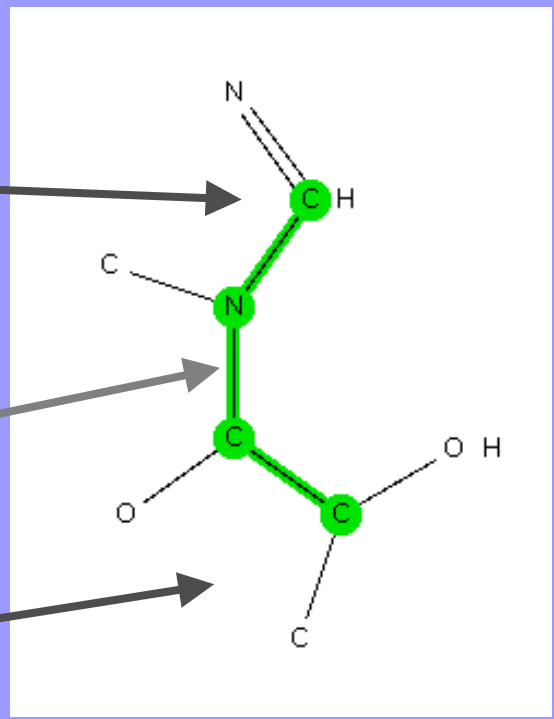
# How does it do that?

## A hierarchical key is built up

- element symbol
- number of bonded atoms
- hydrogen count
- *e.g.* C.3.1

- bond types
- *e.g.* 2.1

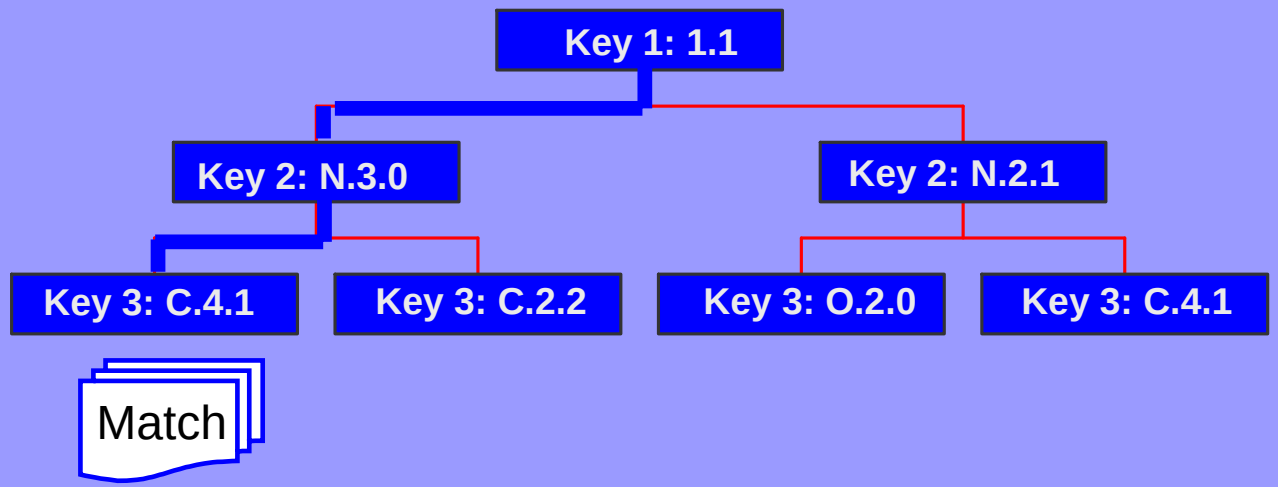
- Ring size *e.g.* 5



Richard Cooper
www.xtl.ox.ac.uk

# How does it do that?

- Search tree created

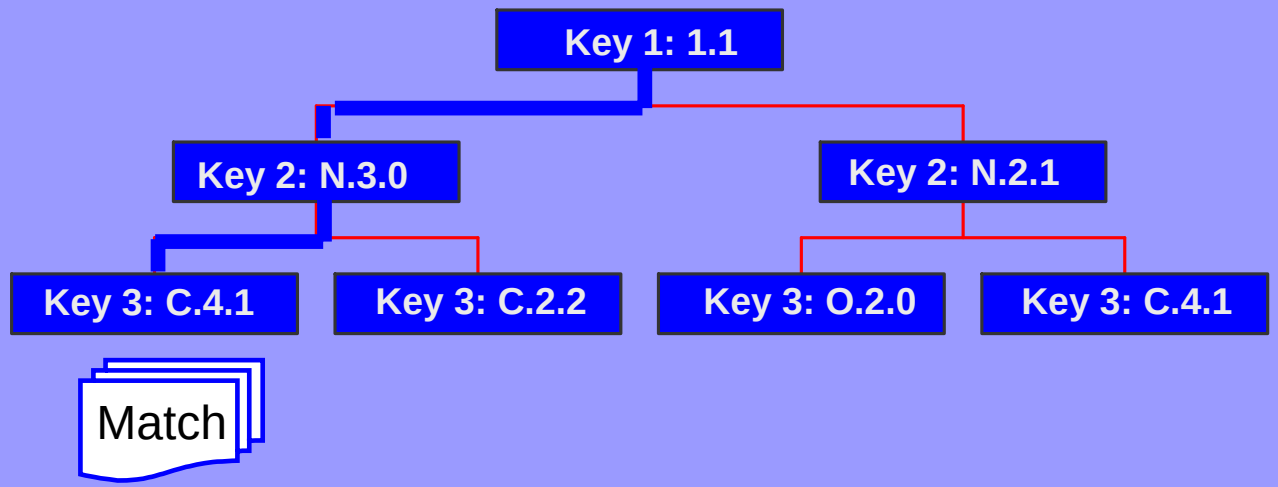



- Each leaf of the tree is searched for in CSD
- When matches are found, statistical quantities are computed and links set to actual structures.

Richard Cooper
www.xtl.ox.ac.uk

# How does it do that?

Keys are assigned to the user's query in the same way, and used to search the tree

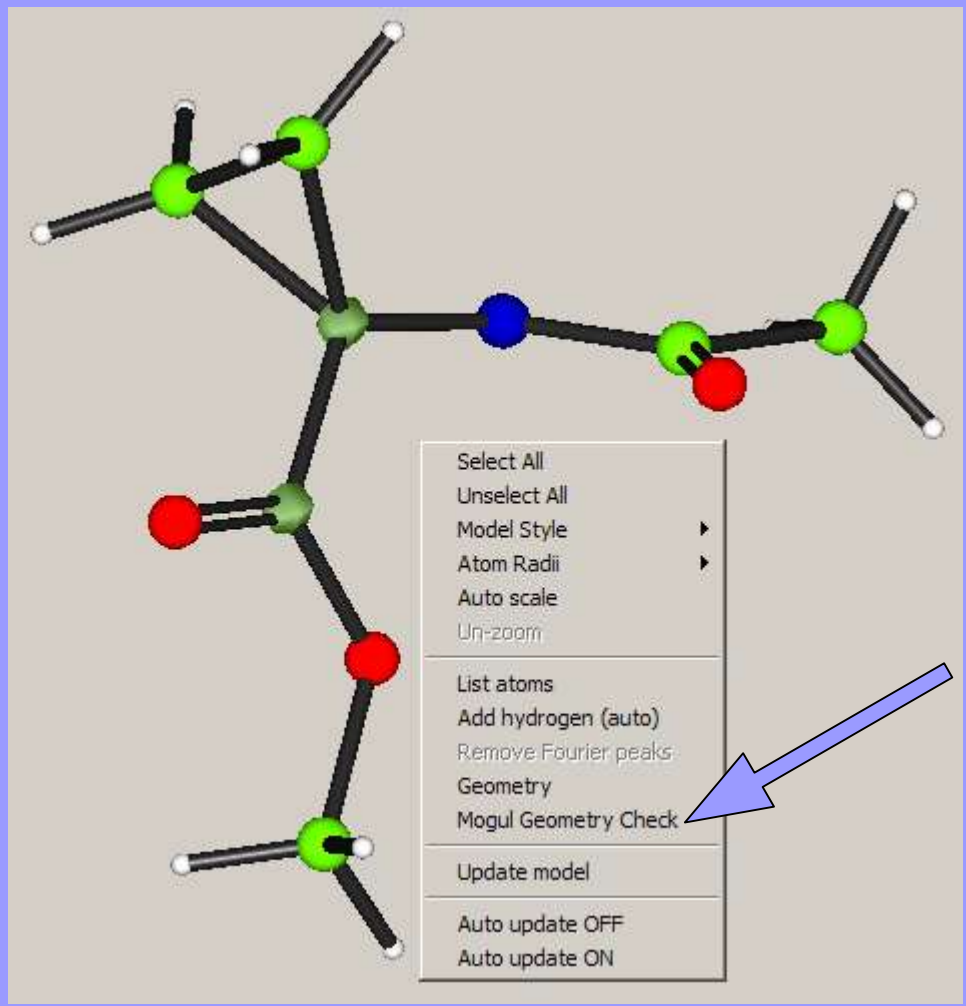



If a match is found, statistical quantities are output and links set to actual structures.

Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

## Simple interaction. MOGUL Geometry Check validates the whole structure

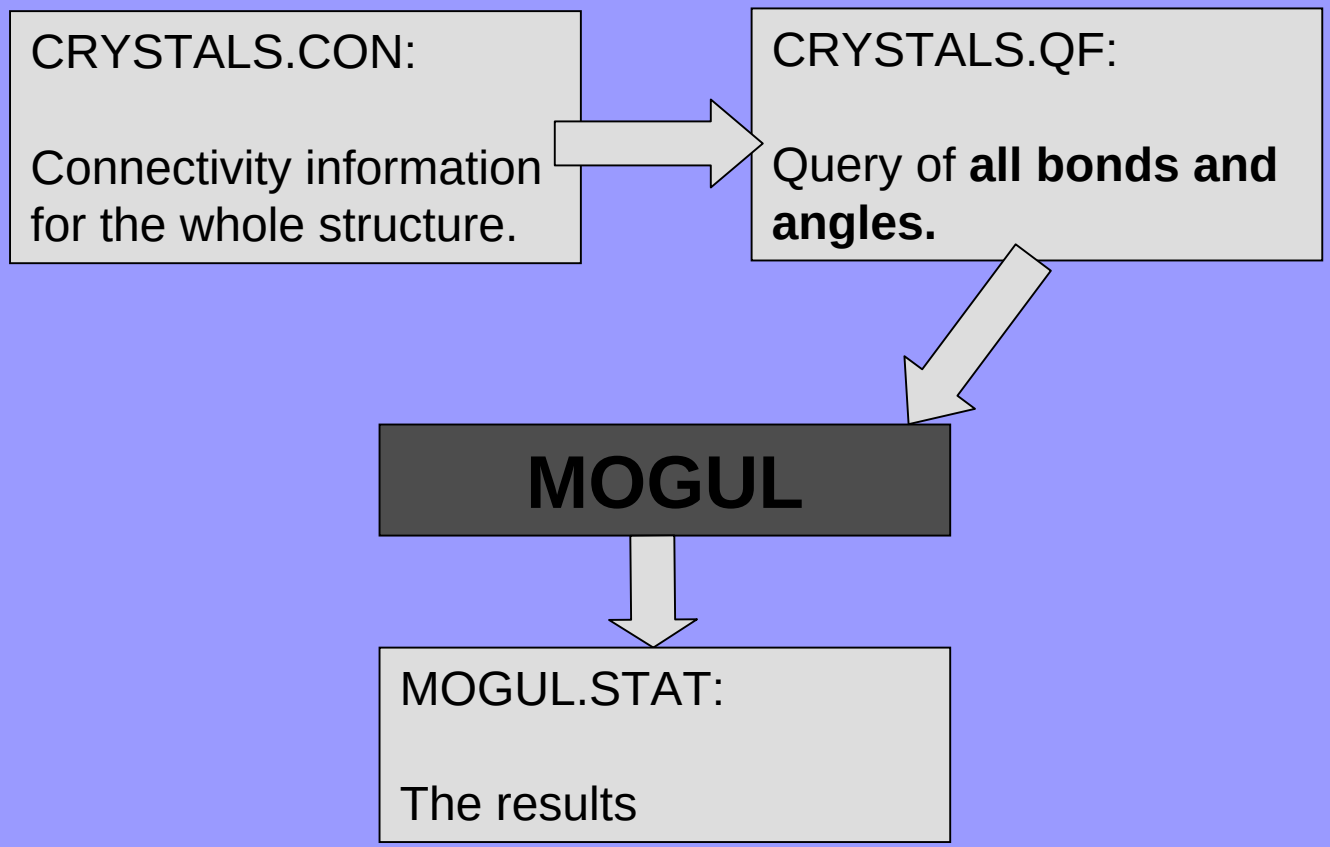


Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

## Automatic interaction:

– Find all unusual bonds and valence angles.



Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

Simple interaction – just click on the menu

MOGUL Geometry Check validates the whole structure

Automatic interaction - find unusual bonds and valence angles.

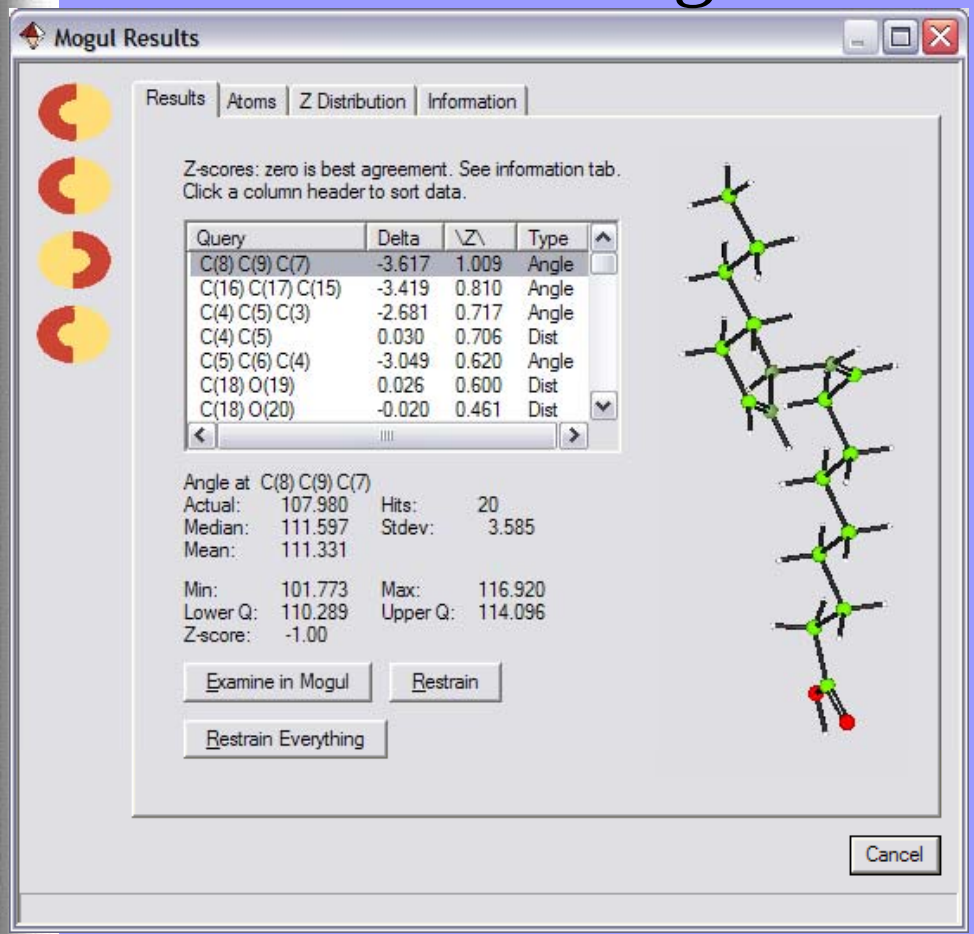
Each bond length and angle is assigned a *z*-score:

$$z = \frac{reported - CSDideal}{\sigma(CSDideal)}$$

Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

Automatic interaction - find unusual bonds and valence angles.




Each bond length and angle is assigned a *z*-score:

$$z = \frac{reported - CSDideal}{\sigma(CSDideal)}$$

Richard Cooper
www.xtl.ox.ac.uk

Mogul Results

Results | Atoms | Z Distribution | Information

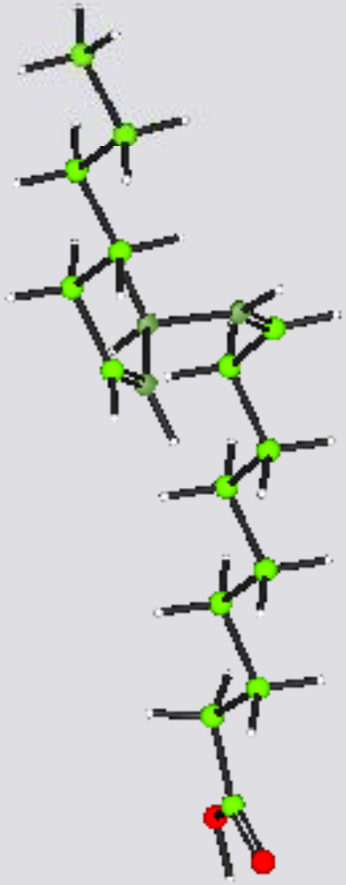
Z-scores: zero is best agreement. See information tab.
Click a column header to sort data.

| Query | Delta | \Z\ | Type |
|---|---|---|---|
| C(8) C(9) C(7) | -3.617 | 1.009 | Angle |
| C(16) C(17) C(15) | -3.419 | 0.810 | Angle |
| C(4) C(5) C(3) | -2.681 | 0.717 | Angle |
| C(4) C(5) | 0.030 | 0.706 | Dist |
| C(5) C(6) C(4) | -3.049 | 0.620 | Angle |
| C(18) O(19) | 0.026 | 0.600 | Dist |
| C(18) O(20) | -0.020 | 0.461 | Dist |

Angle at C(8) C(9) C(7)

| | | | |
|---|---|---|---|
| Actual: | 107.980 | Hits: | 20 |
| Median: | 111.597 | Stdev: | 3.585 |
| Mean: | 111.331 | | |
| Min: | 101.773 | Max: | 116.920 |
| Lower Q: | 110.289 | Upper Q: | 114.096 |
| Z-score: | -1.00 | | |

Examine in Mogul | Restrain | Restrain Everything | Cancel



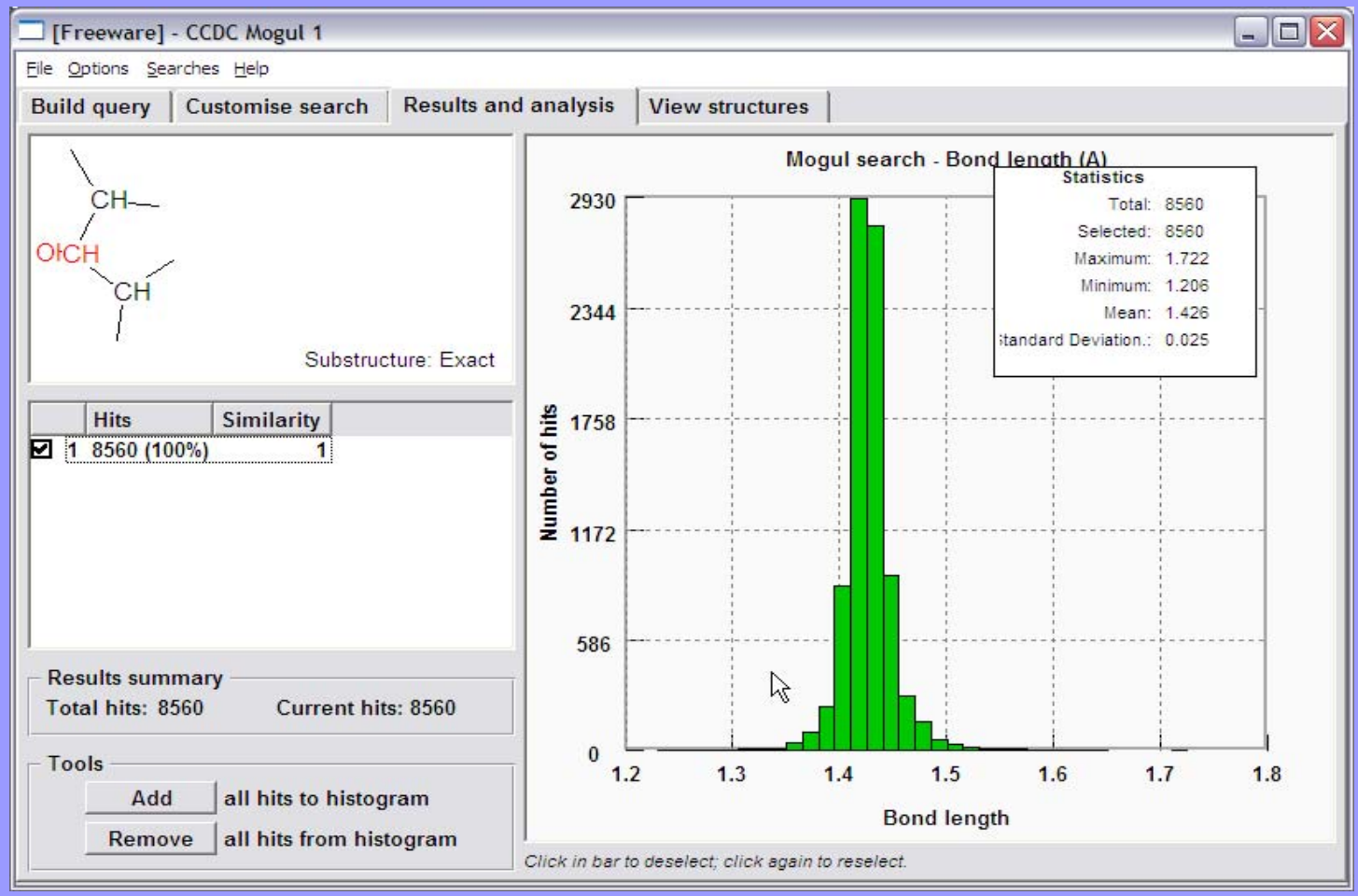
$$z = \frac{reported - CSDideal}{\sigma(CSDideal)}$$

Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

Simple interaction

- Finding out information about an odd looking bond and valence angle.



Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

## Combined figures of merit.

A given atom is generally involved in several bonds and angles, each with an associated z-value.

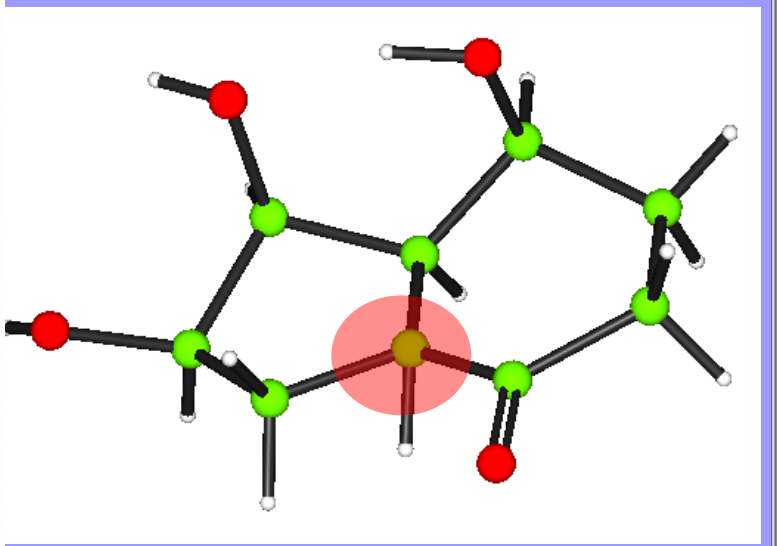
$$FOM = \Pi z_i$$

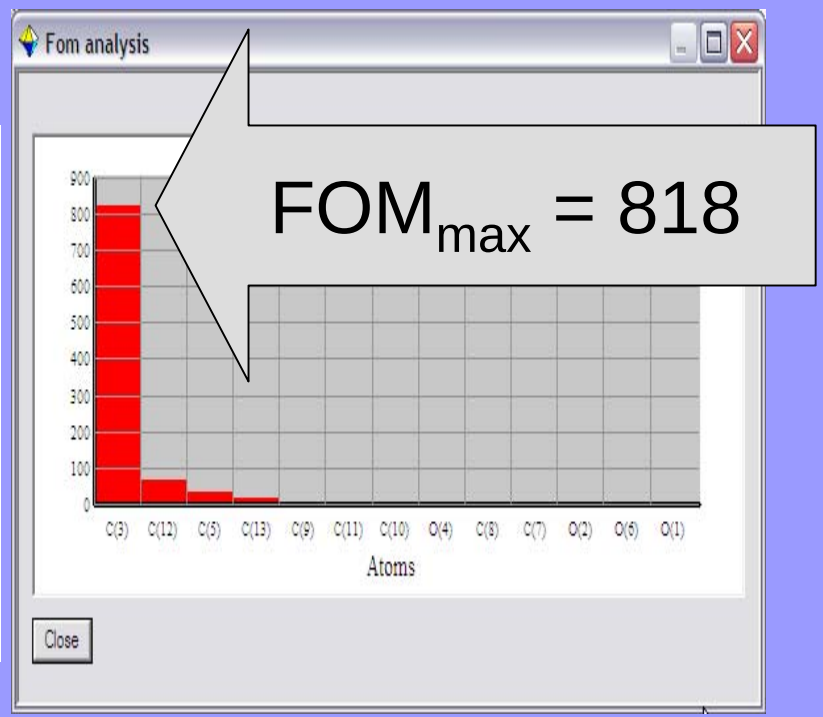
Individual $z_i$ values are limited to the range 1 to 10.

Richard Cooper
www.xtl.ox.ac.uk

# MOGUL and CRYSTALS

## Automatic interaction:






Traps incorrect structures.

Richard Cooper
www.xtl.ox.ac.uk

# What if the fragment does not exist in the CSD?

An interesting observation…

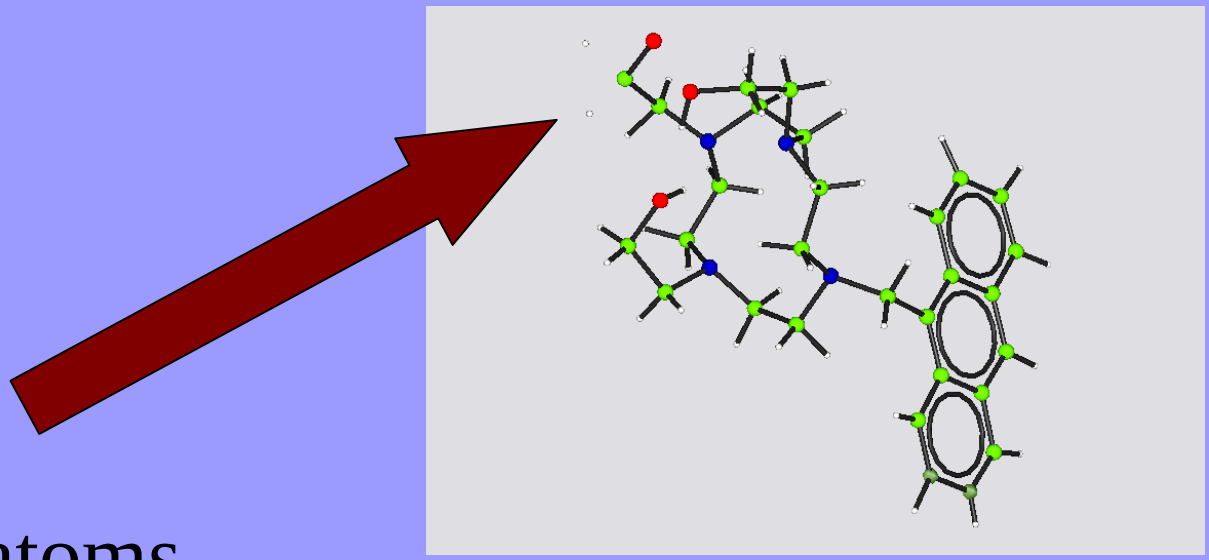
- Decide if the crystallography is secure.
- If it is, publish for the benefit of the community!

# Validating the CSD

Mogul can be used to validate the structures stored in the CSD.

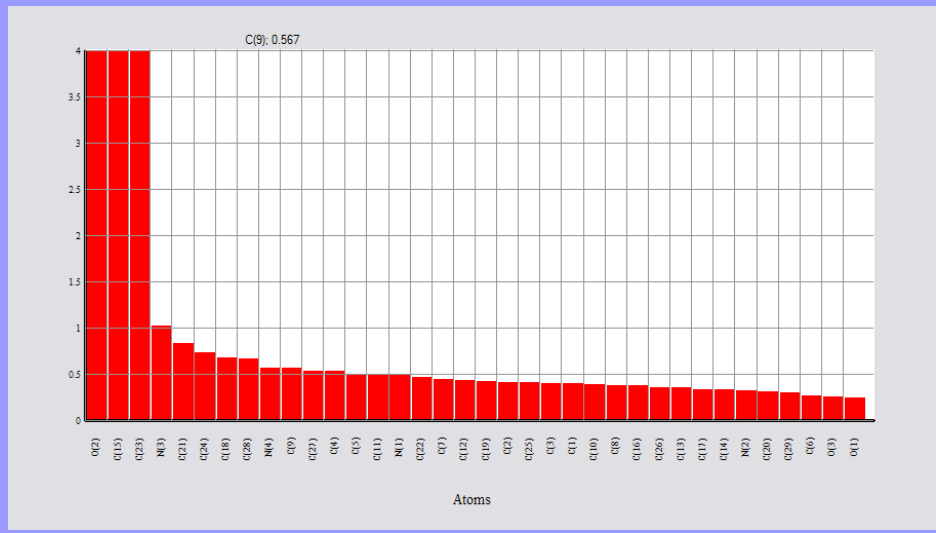
700 structures were extracted from the CSD and run through MOGUL.

The following are examples for which the FOMs were anomalous, possibly pointing to incorrect structures.



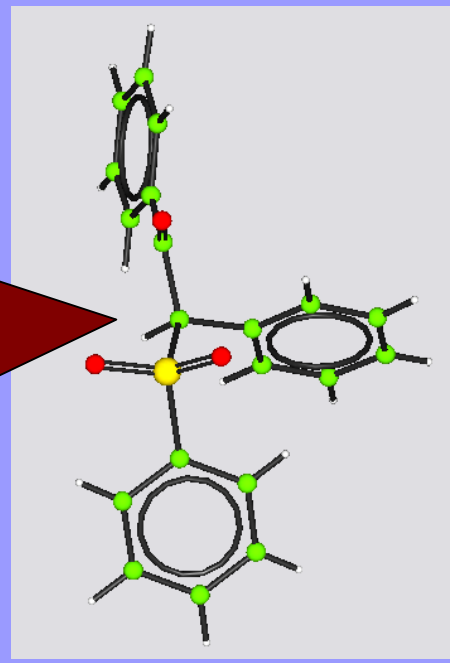
2 hydrogen atoms have fallen off.

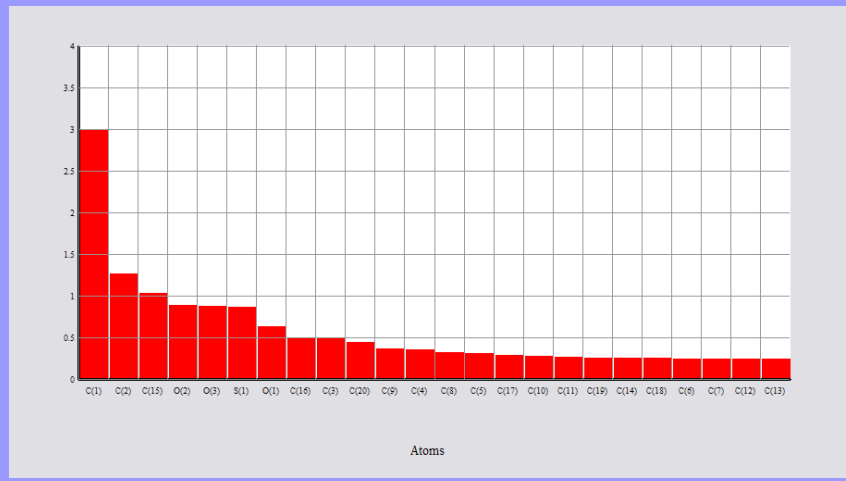
Could be refinement problem, or transcription error.



Very rare environment.

Despite many queries with no results, the FOM fares well – each atom occurs in at least one fragment with a hit.



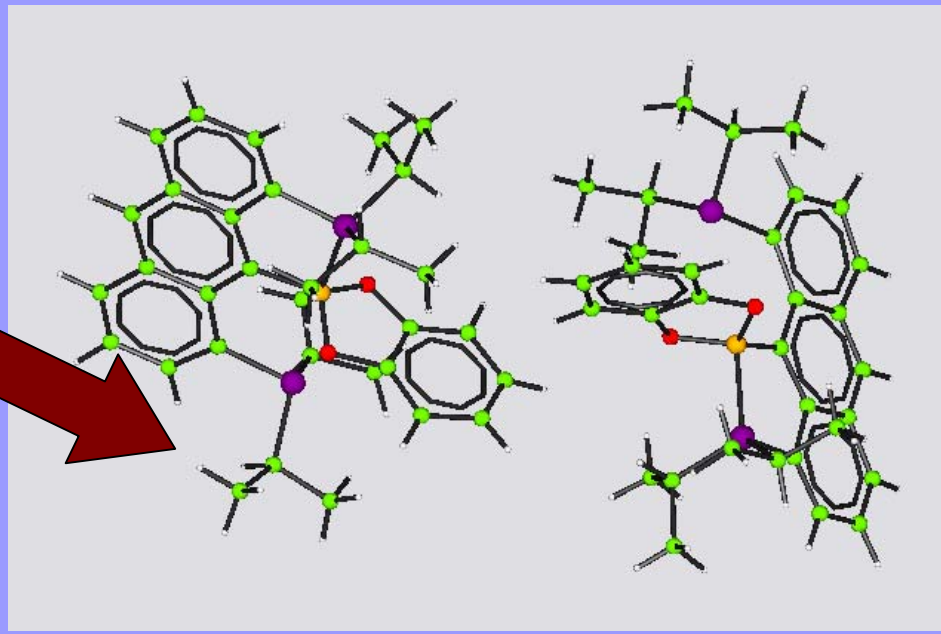


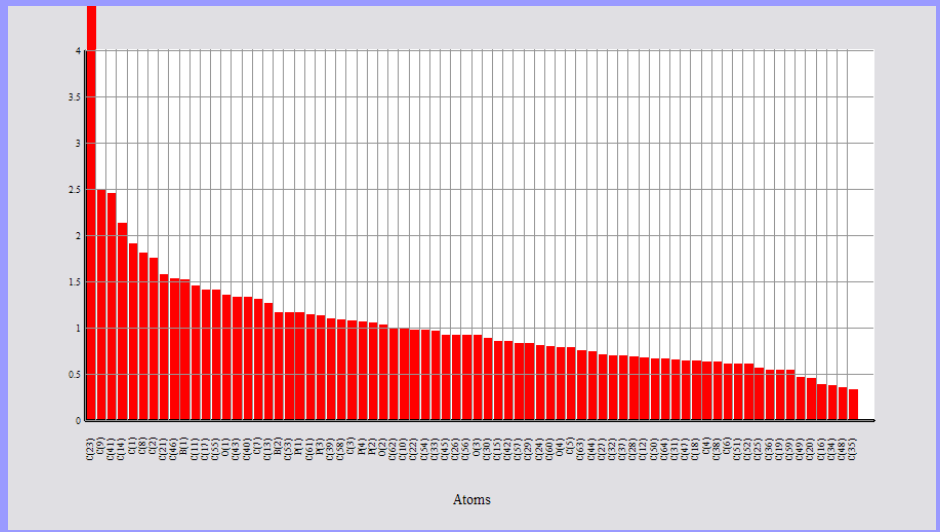
Poor isopropyl groups.

Monoclinic $P2_1$, Z'=2, and $\beta$=90.46°

Pseudo $2_1$ operator along x at y=0, z=0.5 relates molecules.

SG might be $P2_12_12_1$.



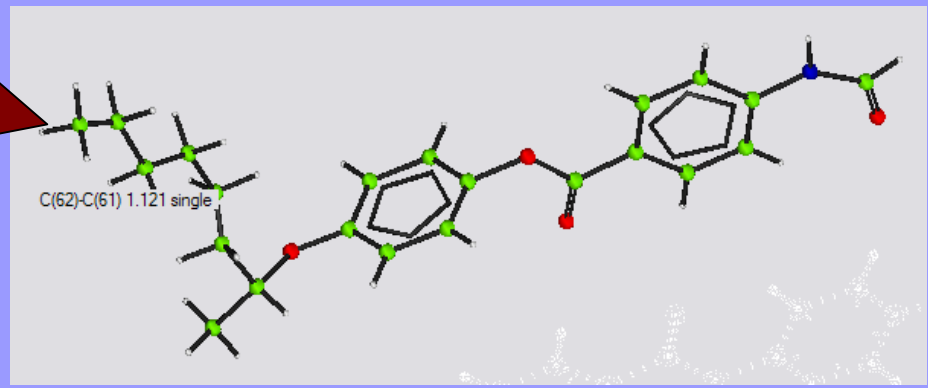


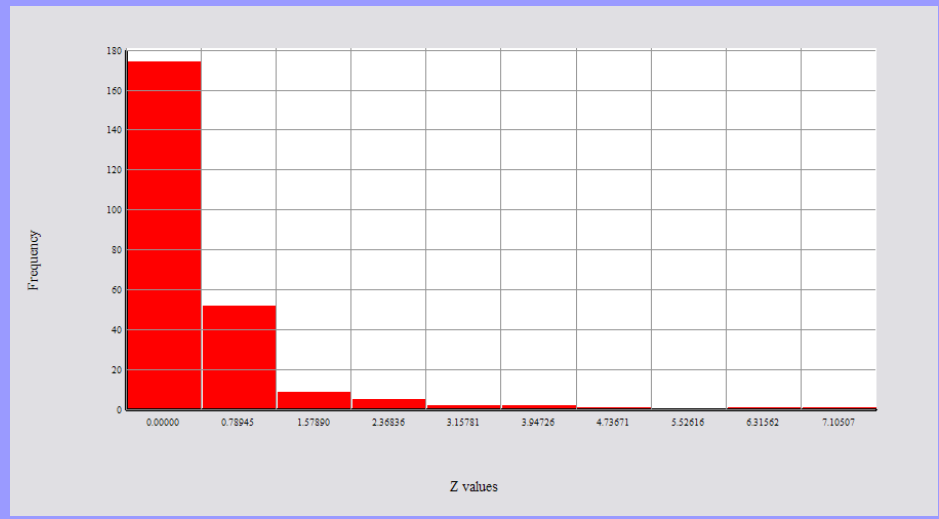
Poor alkyl chains.

Triclinic P1, Z'=4.

$Csp^3$-$Csp^3$ distance of 1.12Å.

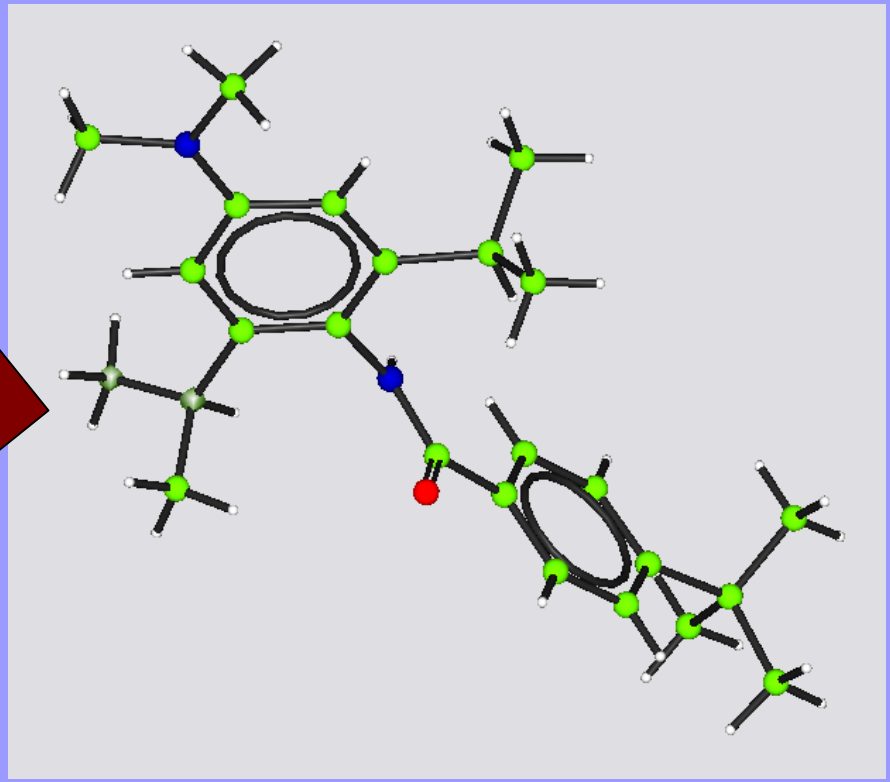
Z distribution tail very long.



Poor isopropyl group. C-C bonds 1.34 Å and 1.45Å [ideal is 1.52Å].

C-C-C angle is 104.3° [ideal is 110.4°]



Z distribution shows these outliers. Possible outliers in $F_o$ data?