Addressing the challenges in applying crystal structure prediction to pharmaceutical materials

PPXRD-10

Lyon, 18 May, 2011

Graeme M. Day
Department of Chemistry
University of Cambridge

http://www-day.ch.cam.ac.uk

This document was presented at PPXRD - Pharmaceutical Powder X-ray Diffraction Symposium

Sponsored by The International Centre for Diffraction Data

PPXRD Website – www.icdd.com/ppxrd

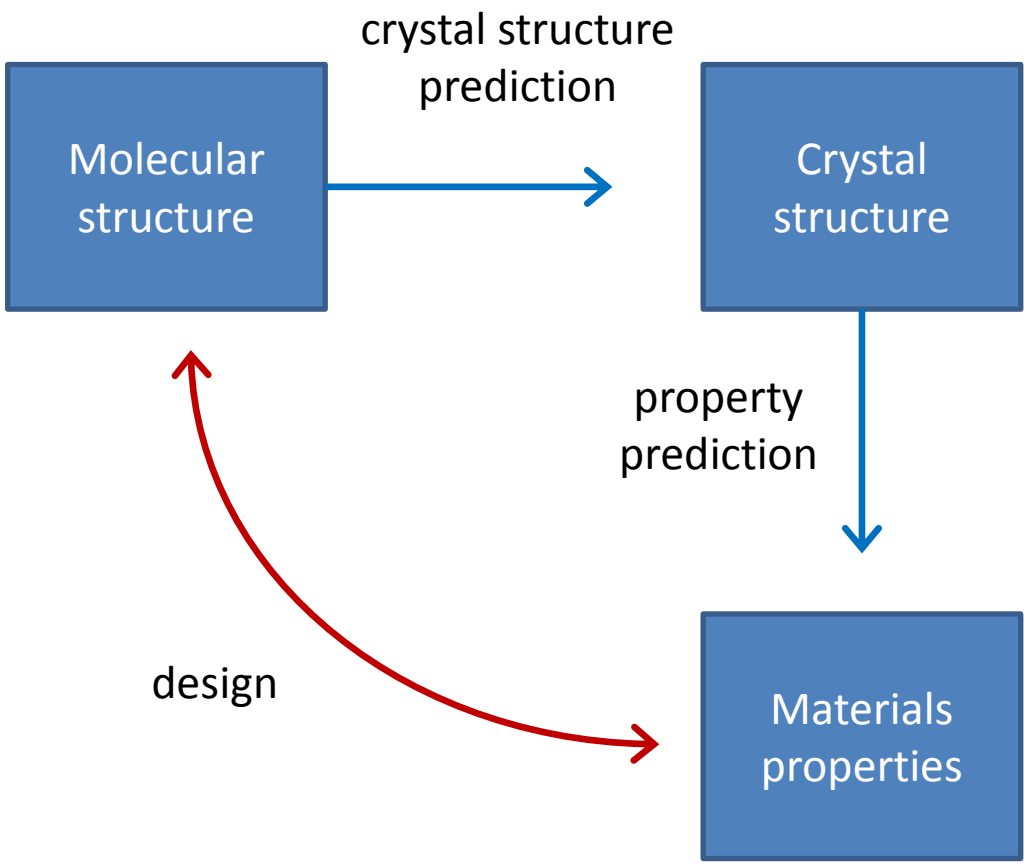
ICDD Website - www.icdd.com

Outline

- crystal structure prediction: aims and methods
- early days of CSP: small, rigid molecules
- solvate / co-crystals
- molecular flexibility

Aim

Development of reliable computational methods for predicting crystal structures and properties.



<u>Why?</u>

Basic understanding
- structure-directing interactions
- structure-property relationships

Polymorphism
- pharmaceutical implications

Crystal engineering / materials design

<u>Property prediction</u>

Some properties of interest:
- solubility (and dissolution rate)
- mechanical properties (tabletability)
- crystal habit (processability)

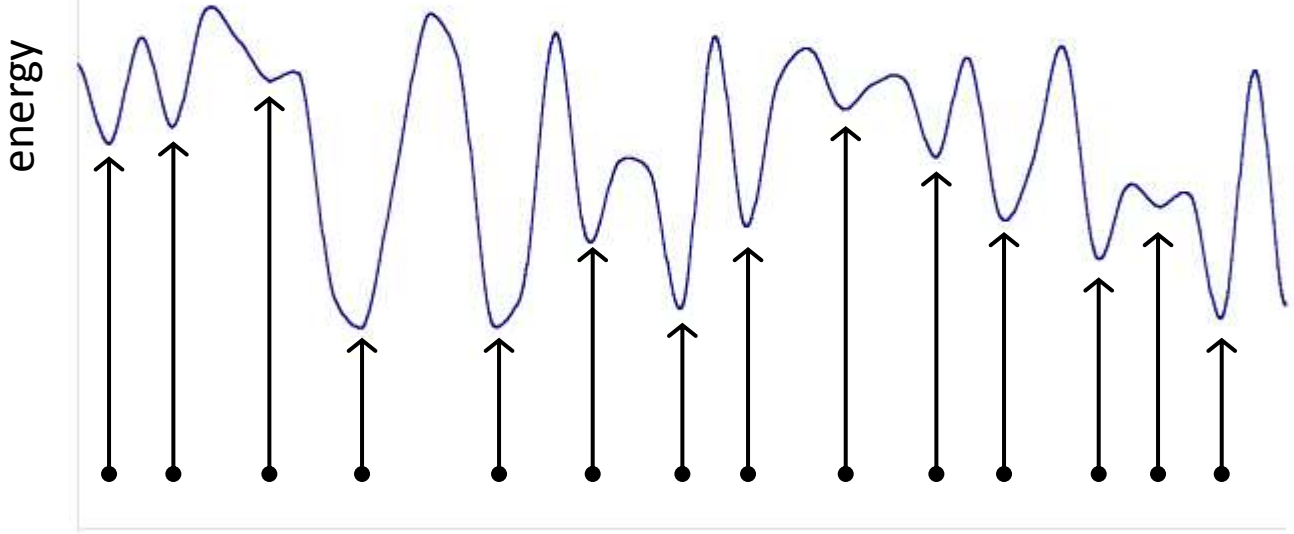
Crystal structure prediction overview of methodology

CSP by global lattice energy minimisation

Two key steps:

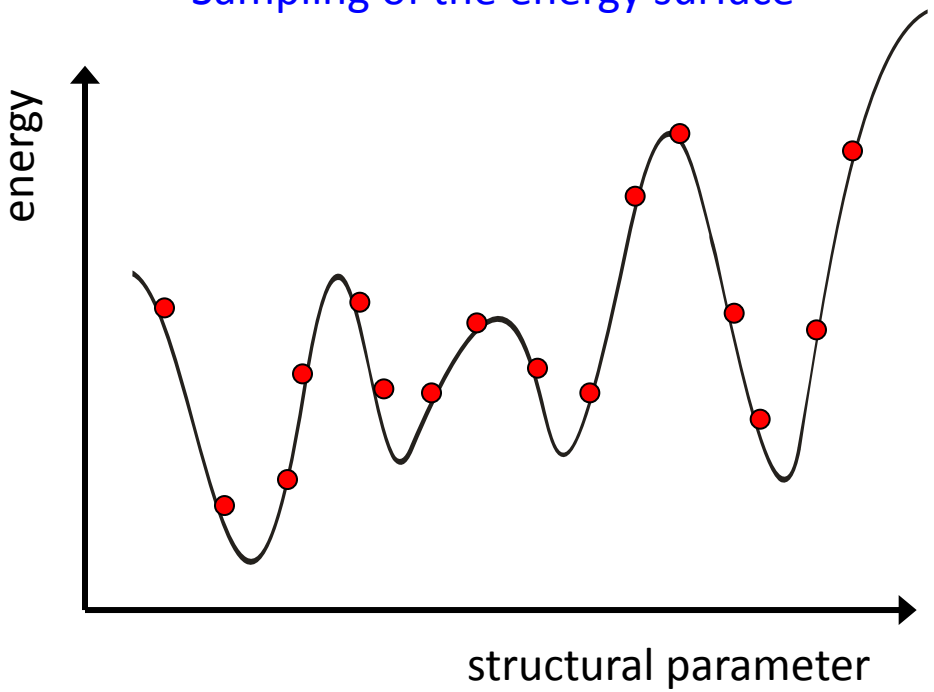
1. explore the potential energy surface
   - all local energy minima are potential crystal structures

2. assess each structure
   - main assumption: lowest energy (global minimum) structure = most likely
   - additional criteria can be added

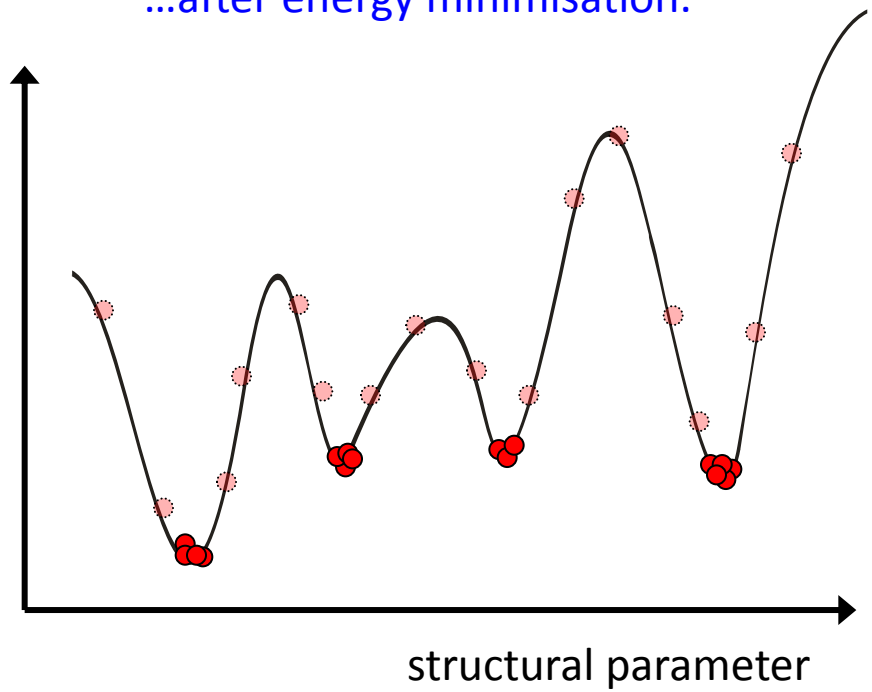


Exploring the lattice energy surface

Sampling of the energy surface



…after energy minimisation.

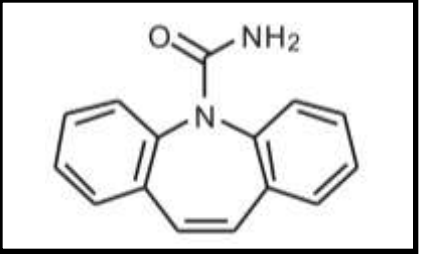


Many algorithms: Monte Carlo, simulated annealing; basin hopping; genetic algorithms; systematic searches; grid, random, quasi-random (low-discrepancy sequences)

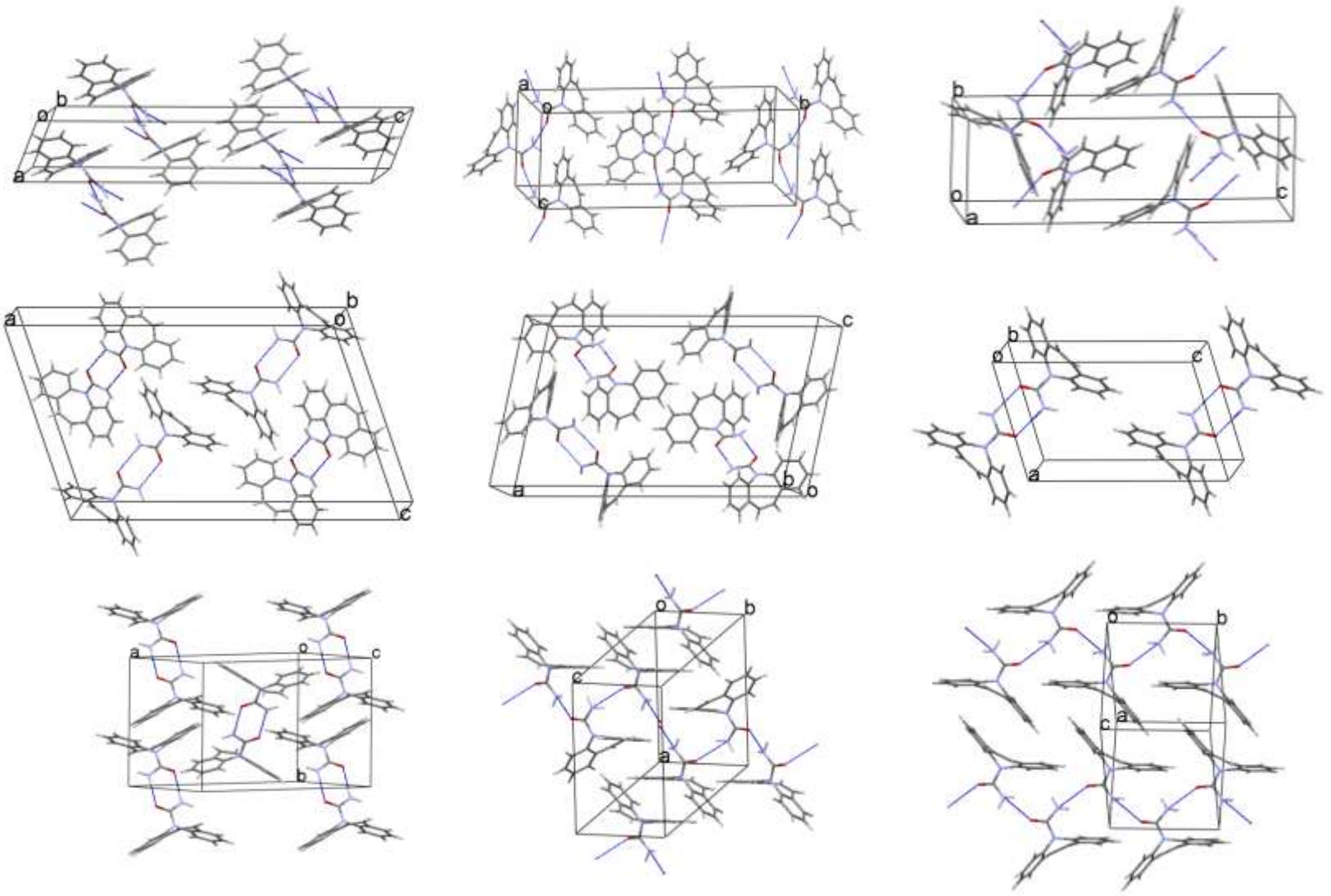
"Clustering": search for and remove duplicate structures

We look to find all low energy minima multiple times.

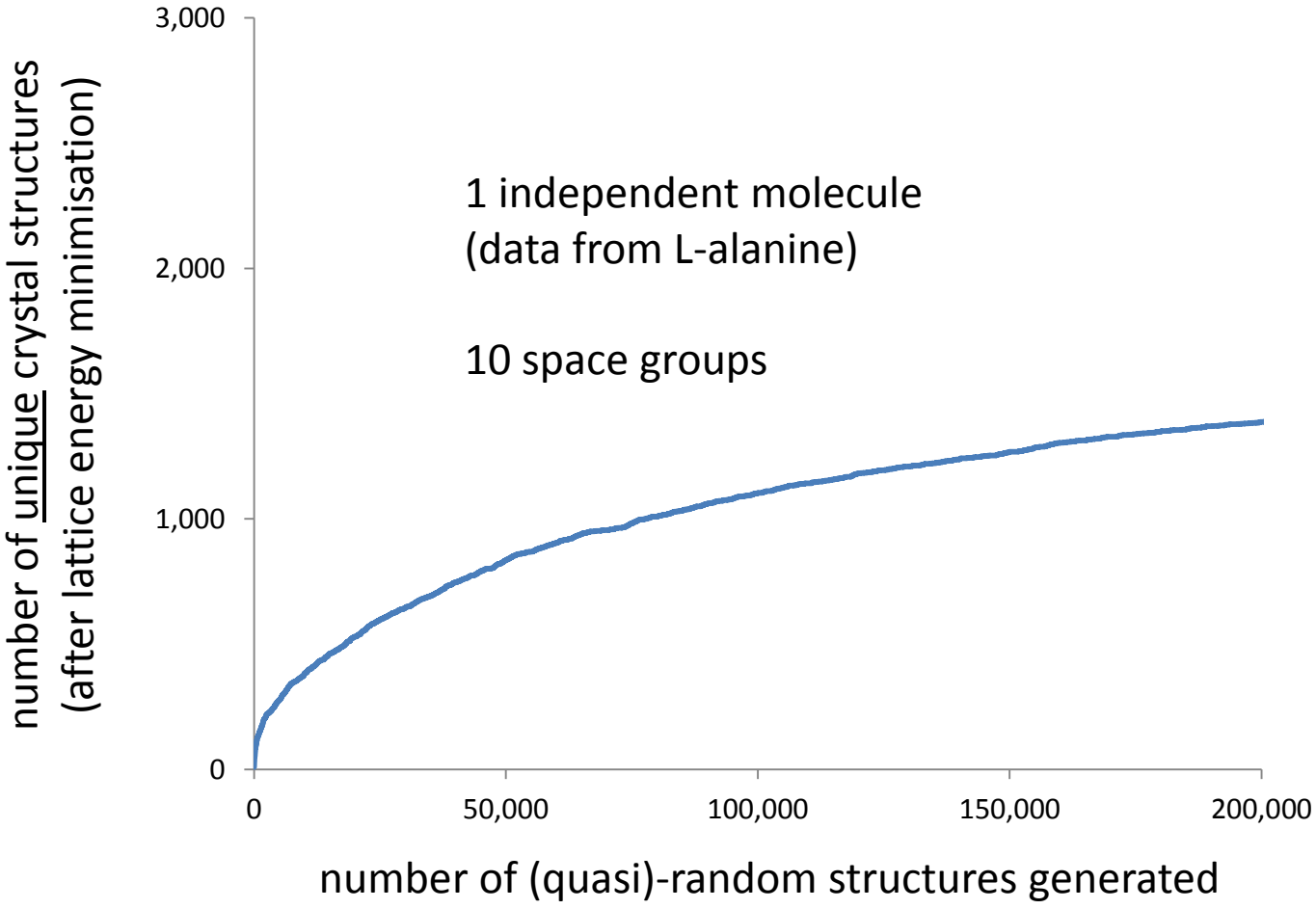
Now we have a set of *distinct* crystal structures and their calculated energies.



structure search + lattice energy minimisation

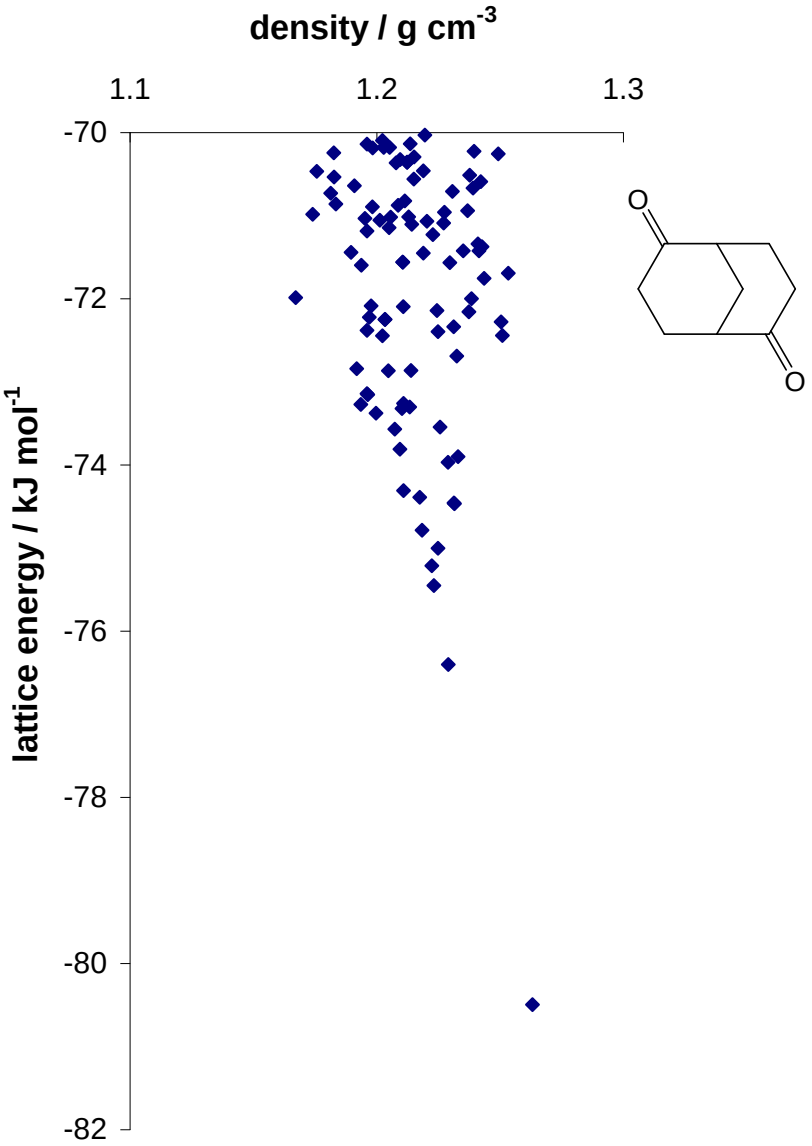


Convergence of a quasi-random structure search

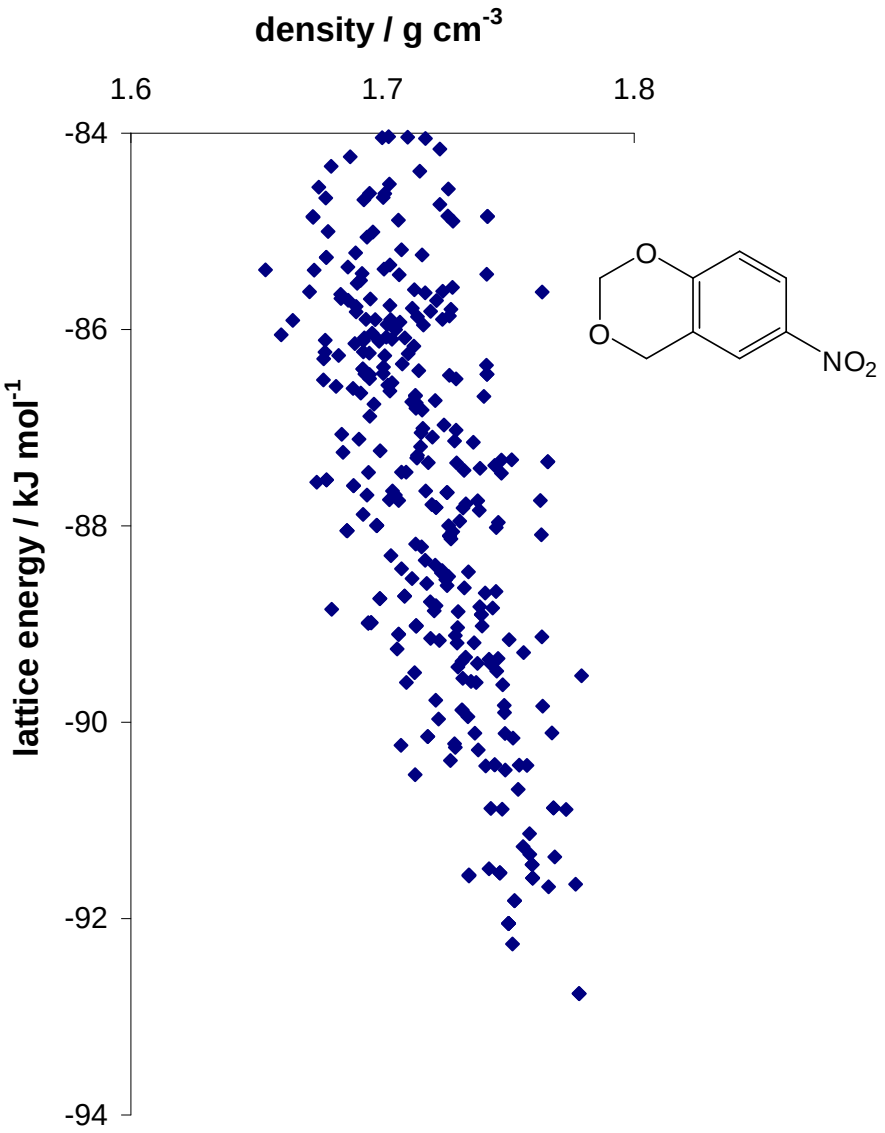


In reality, we must generate and lattice energy minimise 10^5-10^6 trial structures.
Usually leading to ~ 10^4 distinct structures.

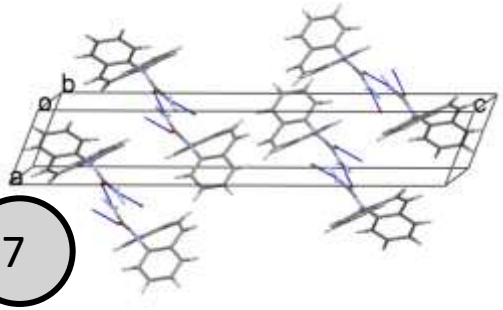
An easily predictable landscape:



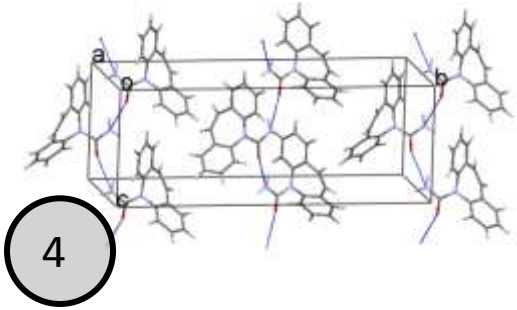
A typical energy landscape:



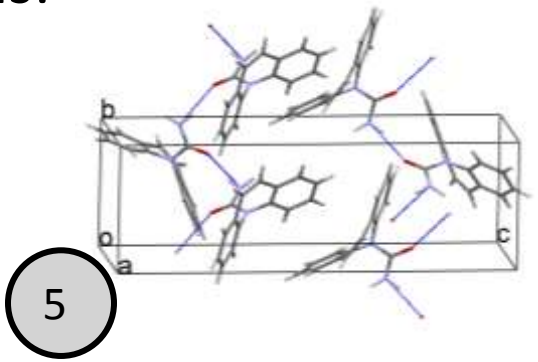
This is a big challenge for computational methods:



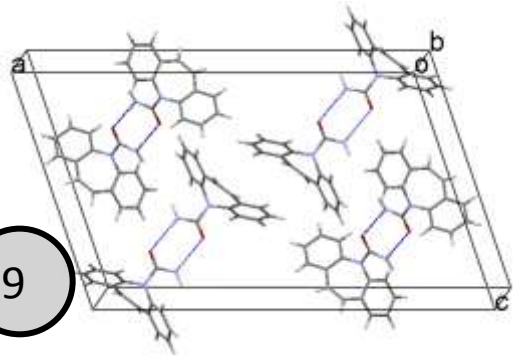
E_{latt} = -112.59 kJ mol^{-1}



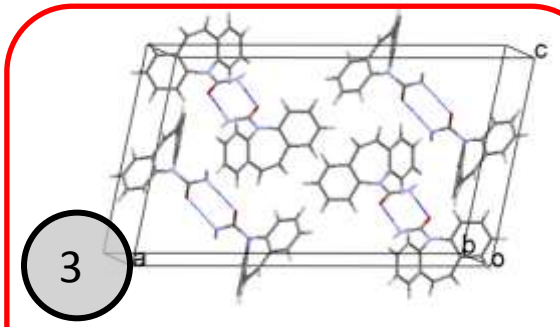
E_{latt} = -113.58 kJ mol^{-1}



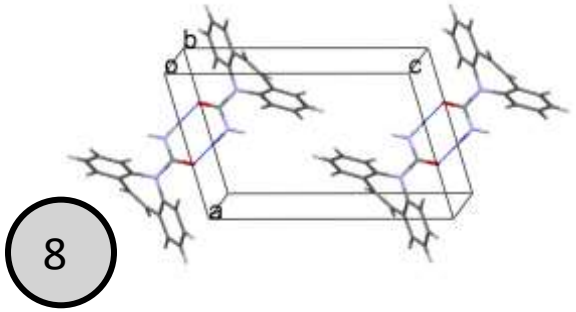
E_{latt} = -113.37 kJ mol^{-1}



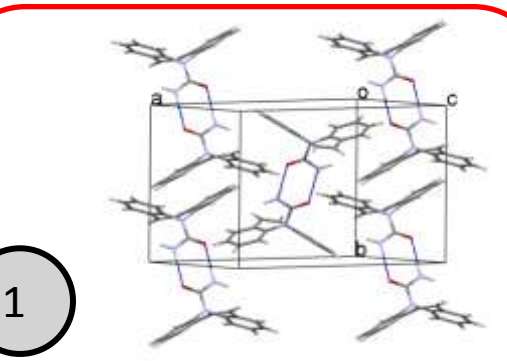
E_{latt} = -111.63 kJ mol^{-1}



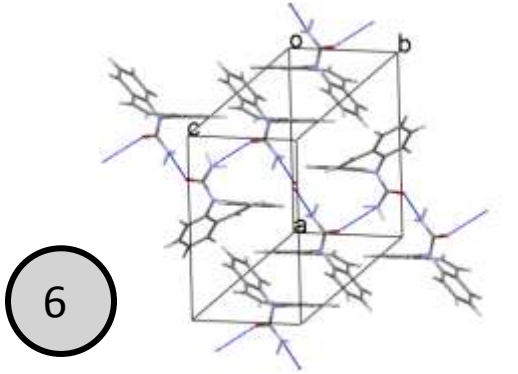
E_{latt} = -114.72 kJ mol^{-1}



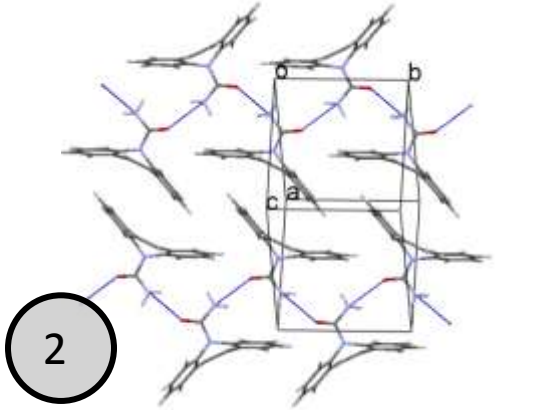
E_{latt} = -112.23 kJ mol^{-1}



E_{latt} = -114.91 kJ mol^{-1}



E_{latt} = -112.69 kJ mol^{-1}



E_{latt} = -114.89 kJ mol^{-1}

approaches to calculating energies:

1) Atom-atom model potentials

Typically, an intermolecular model of the form:

$$U_{ik} = A^{\iota\kappa} \exp\left(-B^{\iota\kappa} R_{ik}\right) - C^{\iota\kappa} R_{ik}^{-6} + U_{electrostatic}$$

$U_{electrostatic}$ comprises either:

a set of atomic partial charges: CPU seconds per crystal structure
or
distributed multipole electrostatic model: CPU minutes per crystal structure

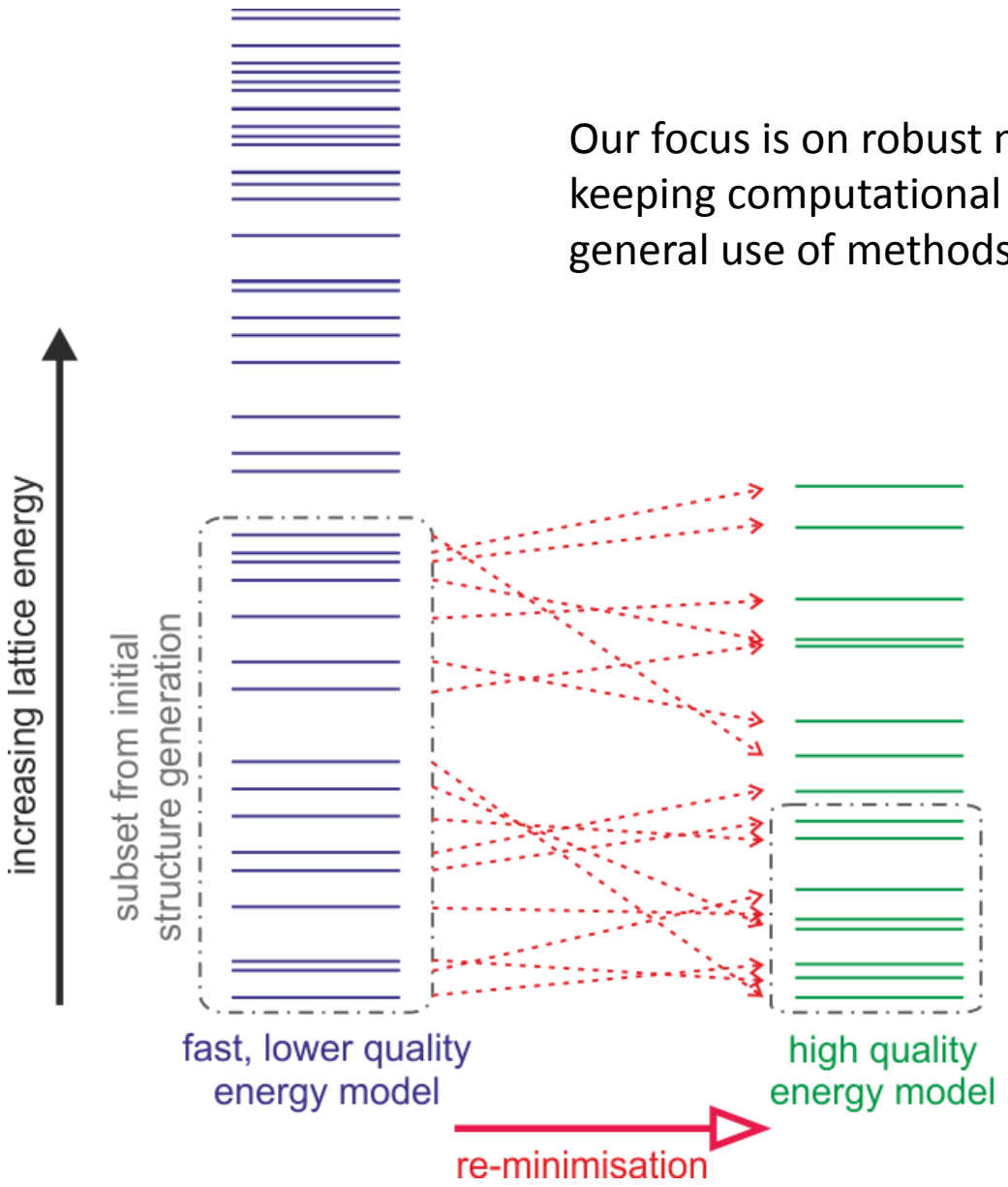
2) Solid state QM methods (DFT, DFT+D)

potentially very accurate

orders of magnitude more expensive: CPU days per crystal structure

Hierarchical approach to structure optimisation and ranking

Our focus is on robust methods, but also on keeping computational expense manageable for general use of methods on a useful timescale.



See Frank Leusen's talk for using electronic structure methods here.

An interesting aside

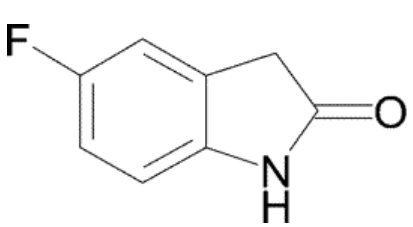
Computers vs People

Do we need accurate energies?

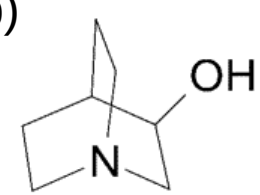
testing intuition / knowledge-based prediction

Can we visually distinguish “good” from “bad” structures?

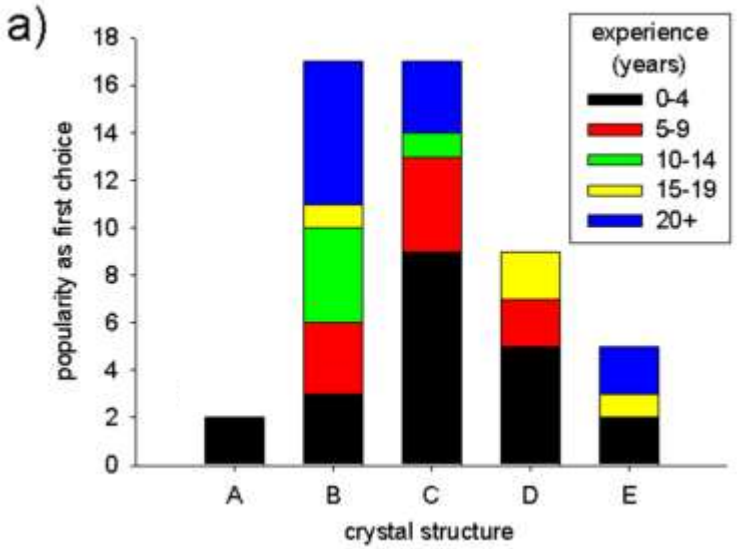
a)

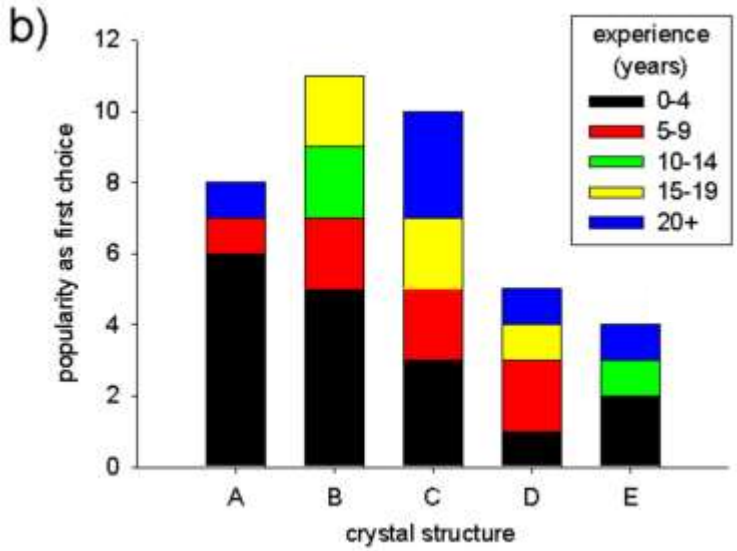


b)

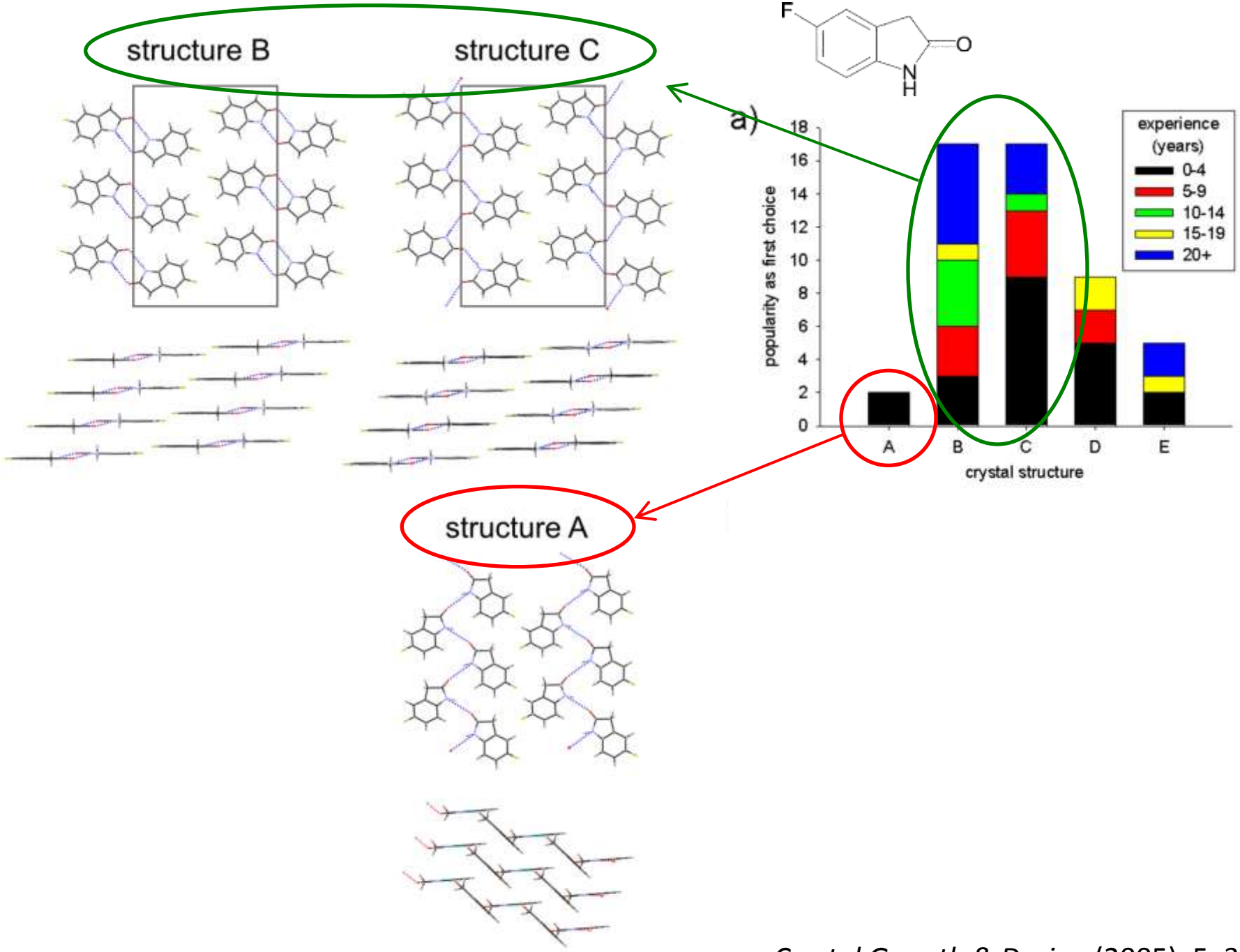


Presented 5 of the lowest energy calculated structures to ~ 50 crystallographers to visually inspect and choose their “favourite”.
(IUCr, Florence, 2005)





Crystal Growth & Design (2005), 5, 391.



Crystal Growth & Design (2005), 5, 391.

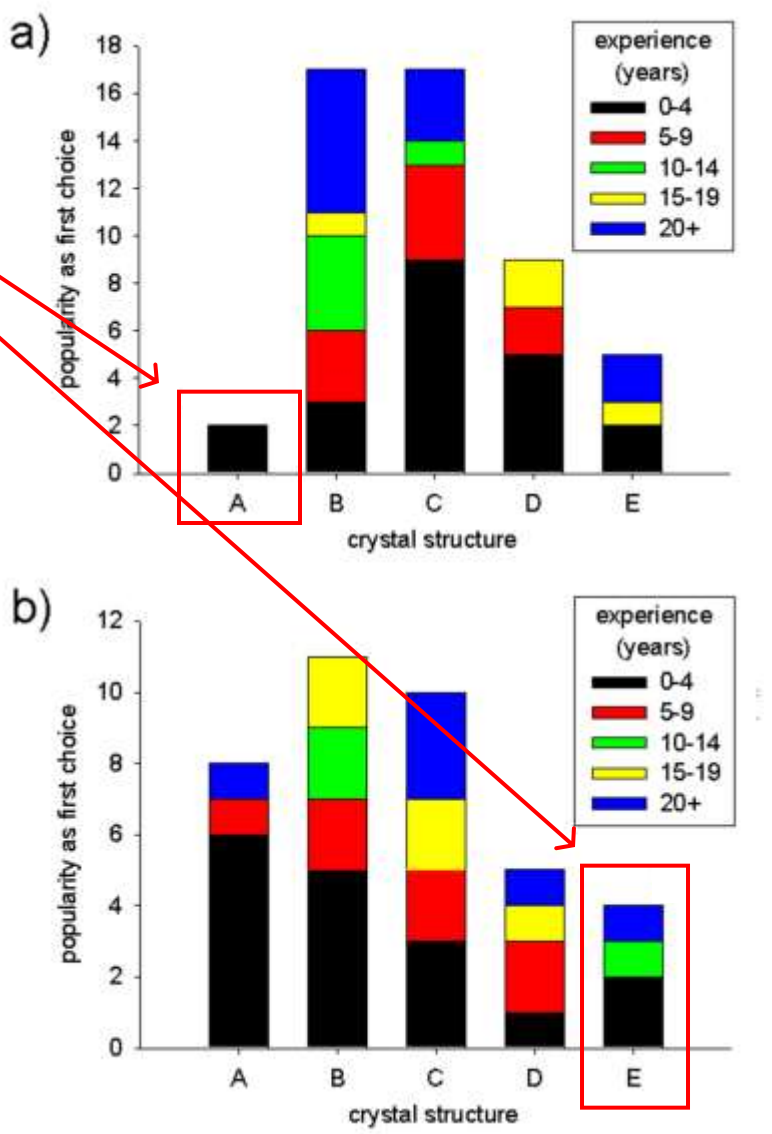
• The observed structure was the *least* preferred in both cases.

• the real structures do “look good”…
…but so do the other predicted structures.
They sometimes even look better.

Lessons:

• intuition can point in the wrong direction

• let’s keep going with energies



Crystal Growth & Design (2005), 5, 391.

Small, rigid molecules

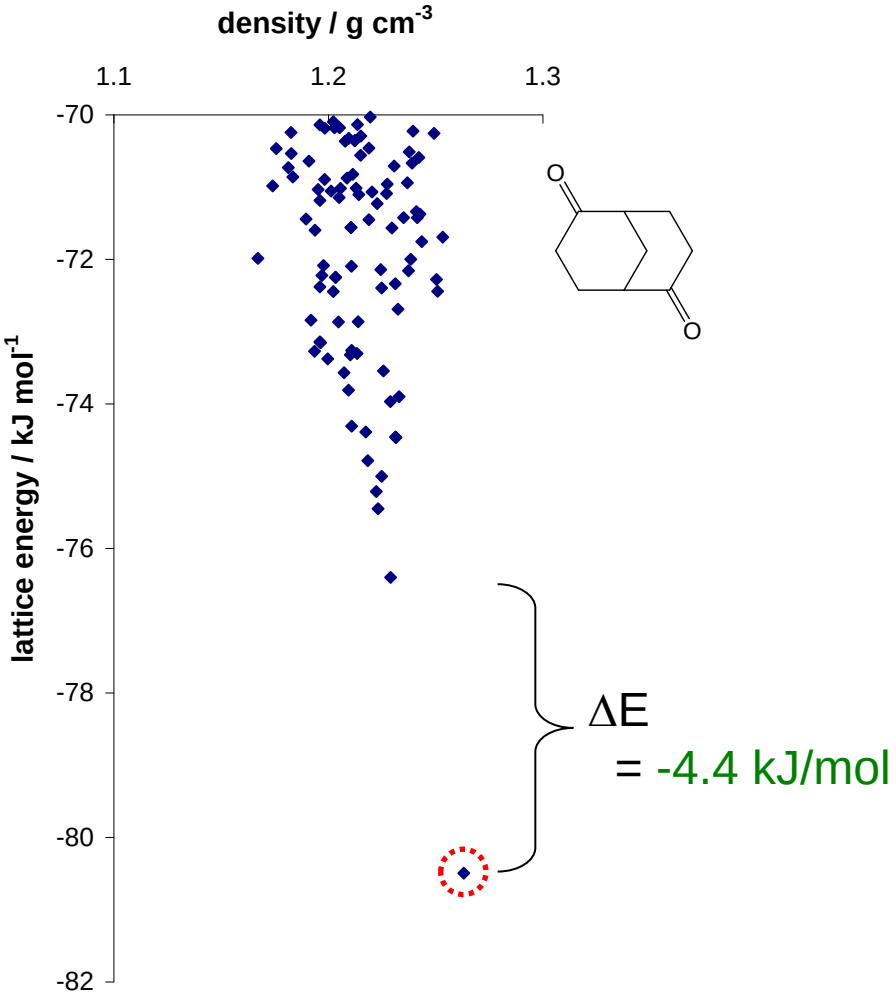
Molecular geometry is assumed unaffected by crystal packing
→ simplifies crystal structure search
→ a test of models of intermolecular interactions

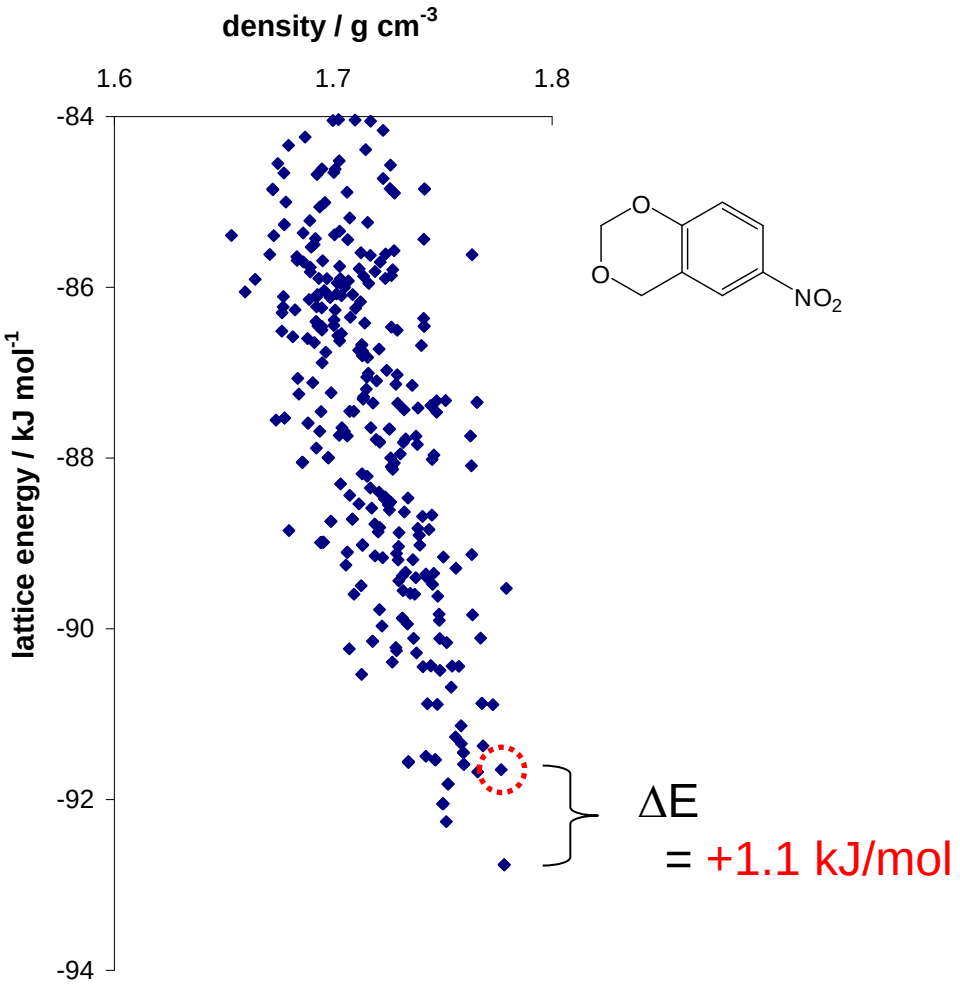
Calculations are fairly fast (days per molecule on 1 CPU)
→ we can look at a large set of molecules
→ assess the global energy minimisation approach

Testing intermolecular models

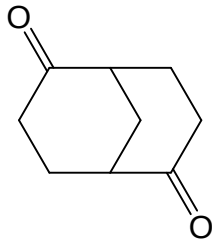
Use ΔE as a measure of success: how far in energy is the "real" structure from the lowest energy predicted structure.

$\Delta E = U_{latt}$(observed structure) – U_{latt}(lowest energy unobserved structure)



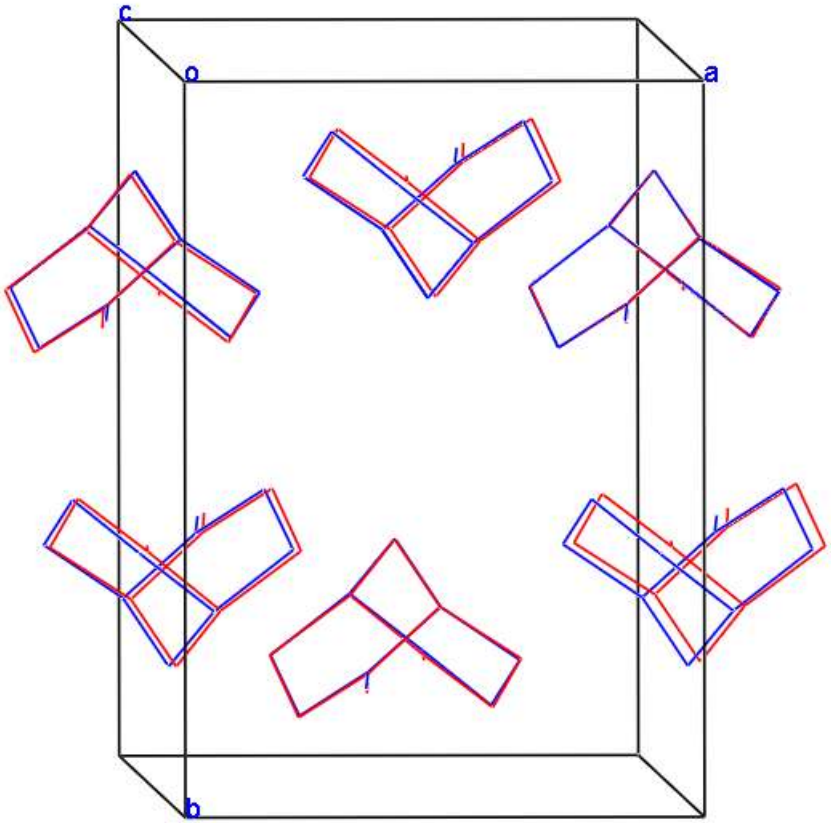


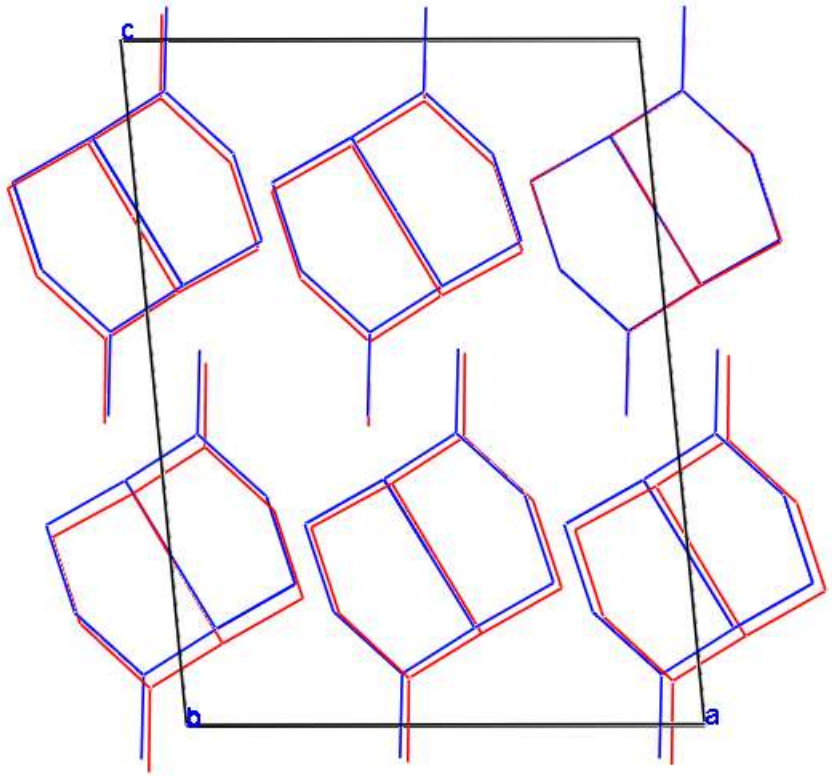
Level of agreement that we aim for:



blue = observed structure (XRD)
red = global minimum predicted structure

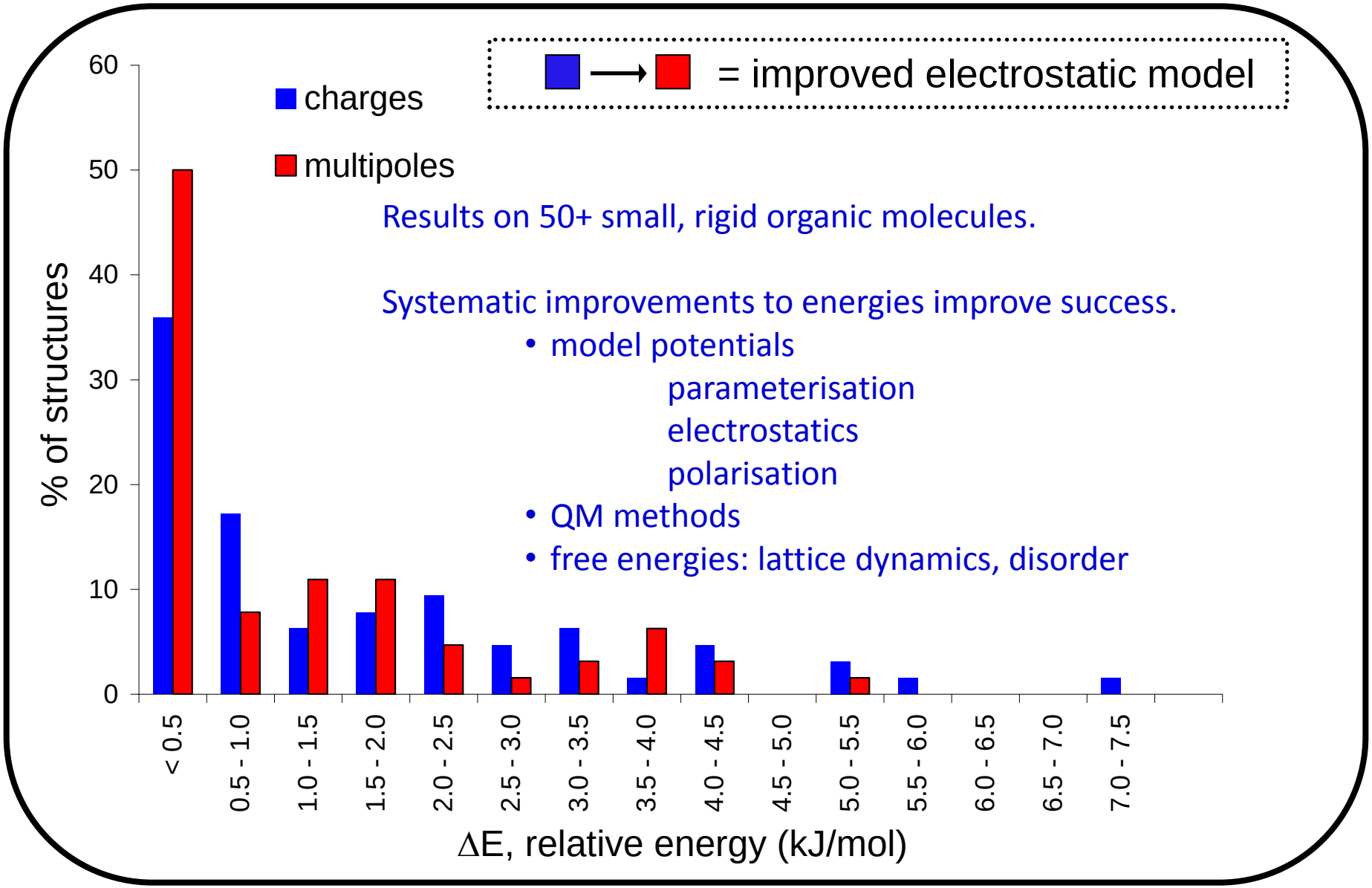
(hydrogen atoms hidden for clarity)





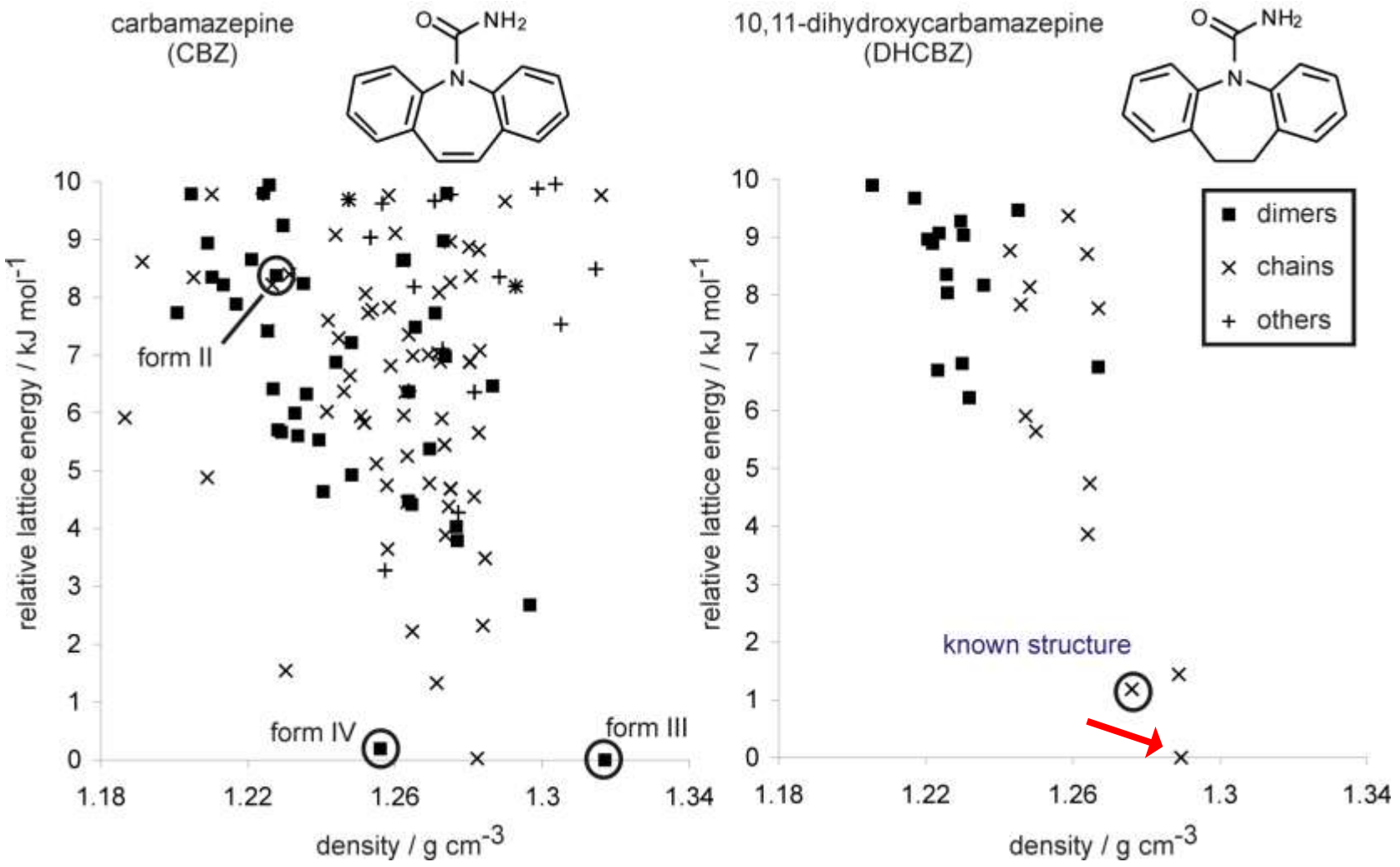
Typical errors are up to 3% in lattice parameters.

Overall results and dependence on the intermolecular potential



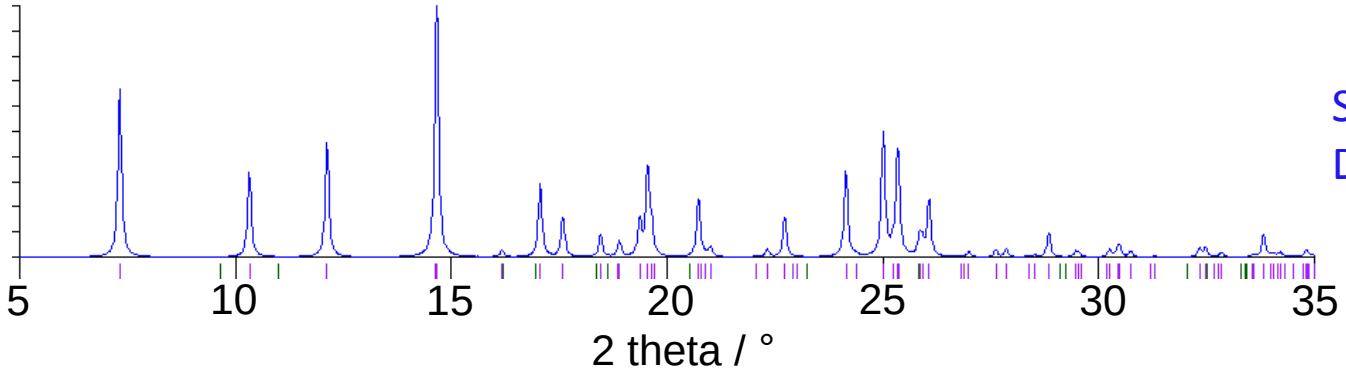
Crystal Growth & Design (2004), 4, 1327; *Crystal Growth & Design* (2005), 5, 1023.

Guiding the experimental discovery of new polymorphs

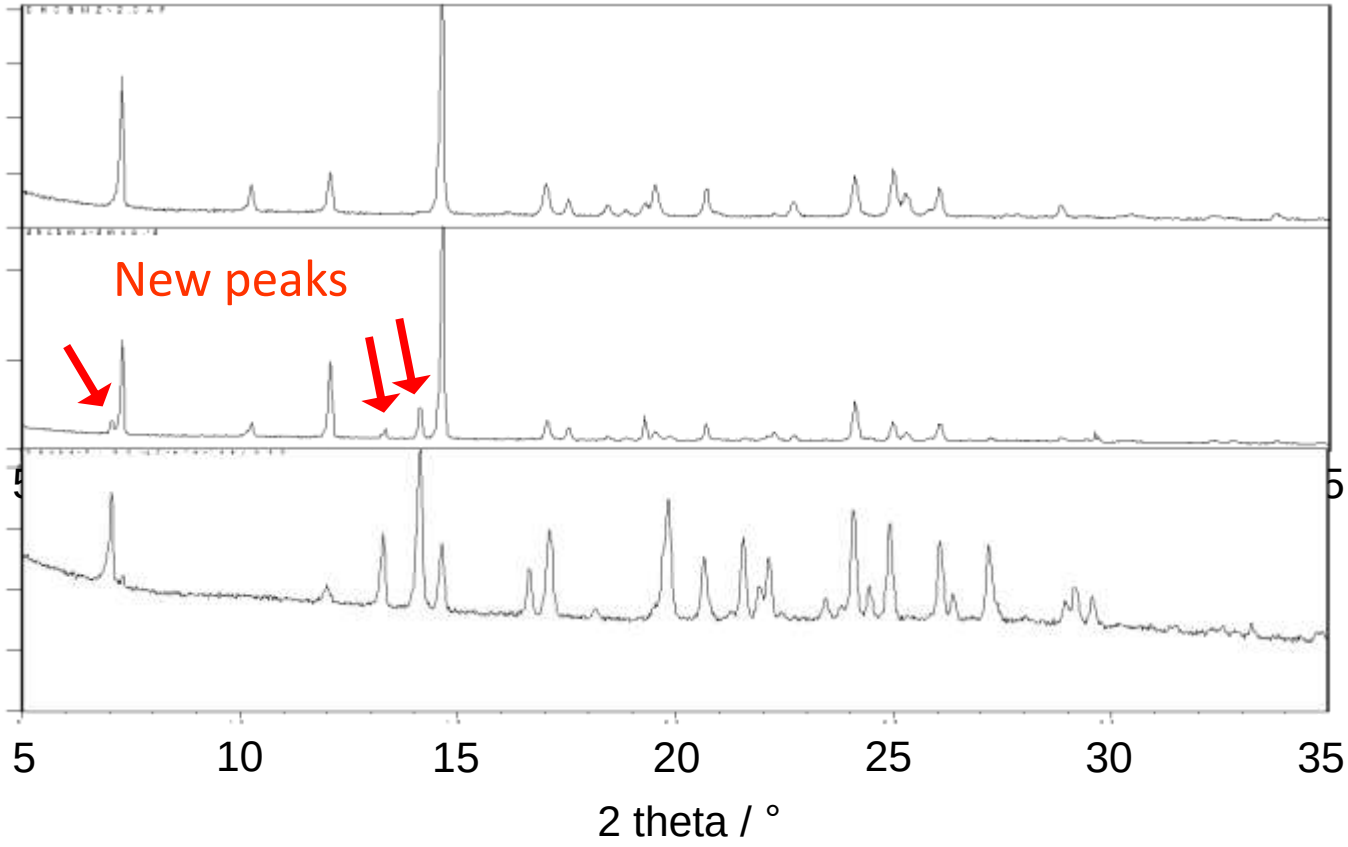


Crystal Growth & Des. (2007), 7, 100-107.

To the lab… varying crystallisation conditions



Simulated from known DHCBZ structure



DHCBZ grown from ethanol

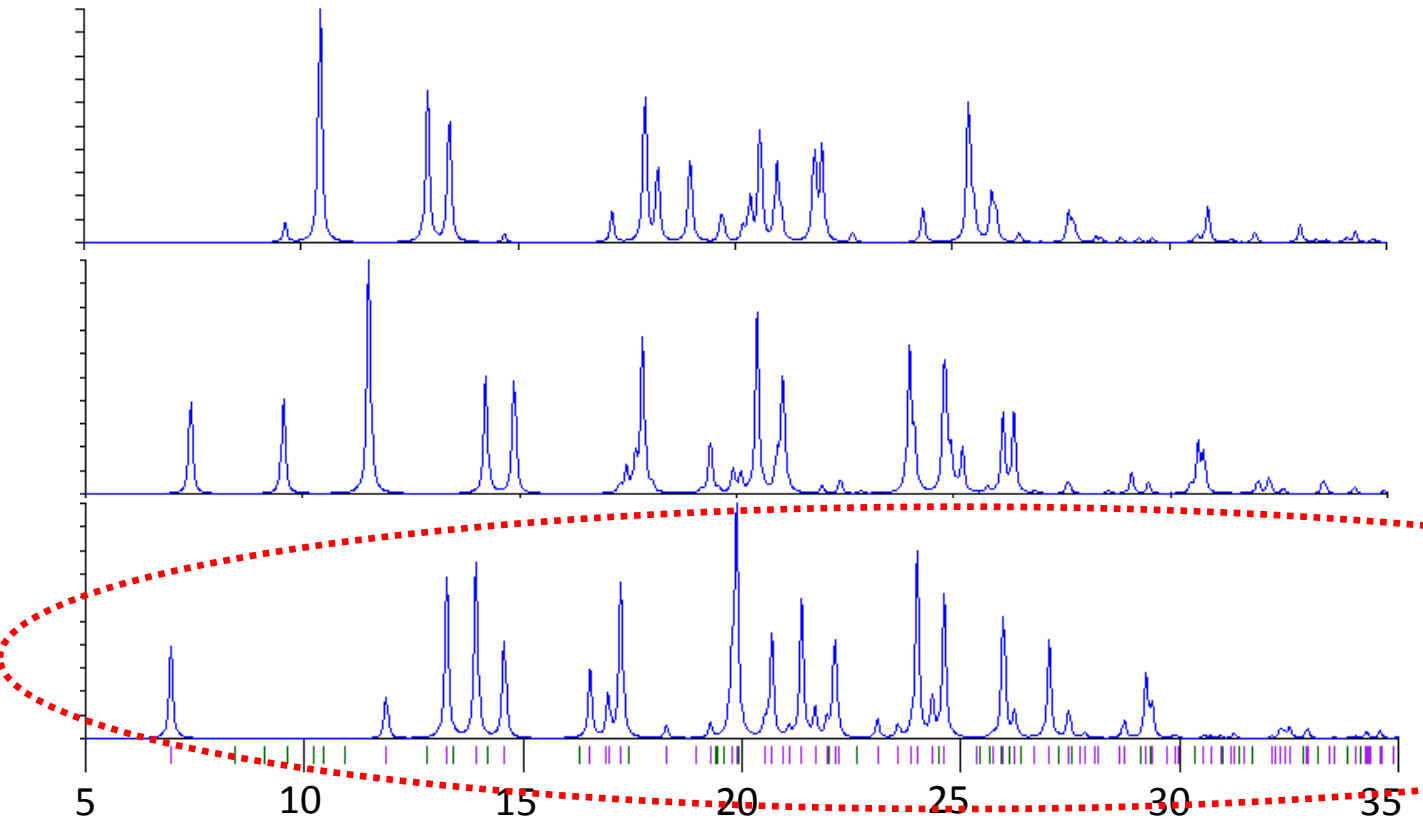
DHCBZ grown from DMSO

DHCBZ ground in DMSO

New form

Crystal Growth & Des. (2007), 7, 100-107.

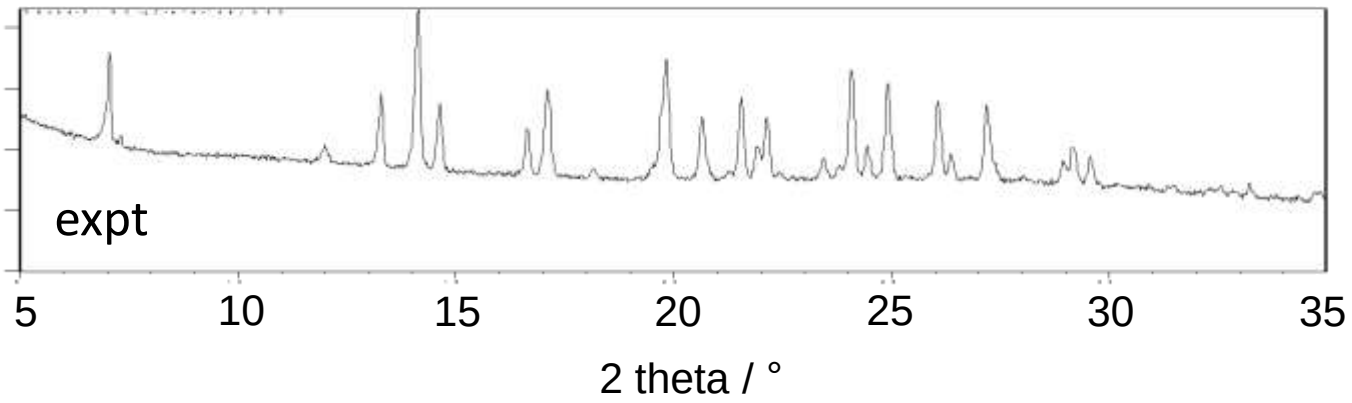
Simulated XRPD from predicted structures



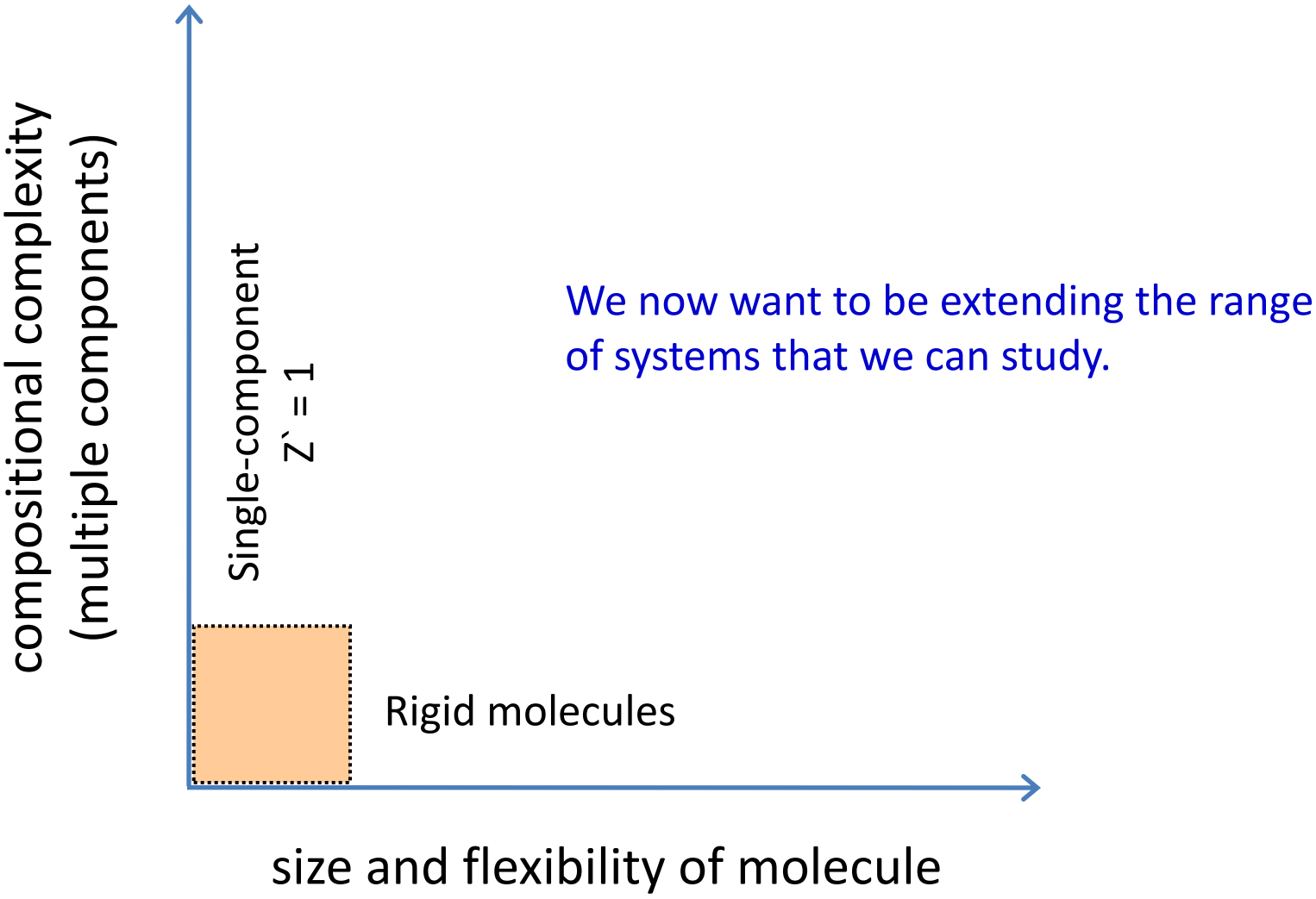
Structure #8
E_{latt} = -112.8 kJ mol^{-1}

Structure #3
E_{latt} = -117.5 kJ mol^{-1}

Structure #1
E_{latt} = -119.0 kJ mol^{-1}



DHCBZ ground
in DMSO
New form



We now want to be extending the range of systems that we can study.

Co-crystallisation & solvate formation

Introducing a second molecular component can tailor properties

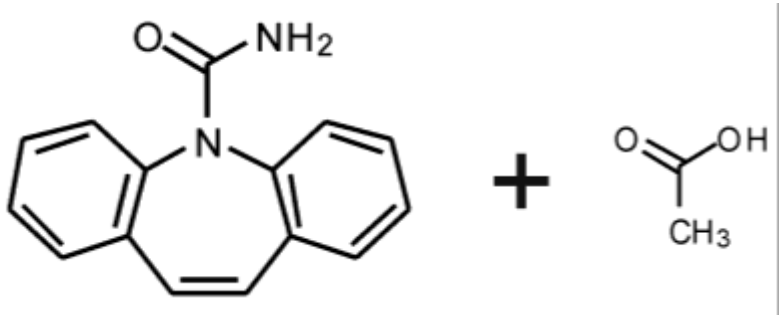
eg. paracetamol (poor compressibility)

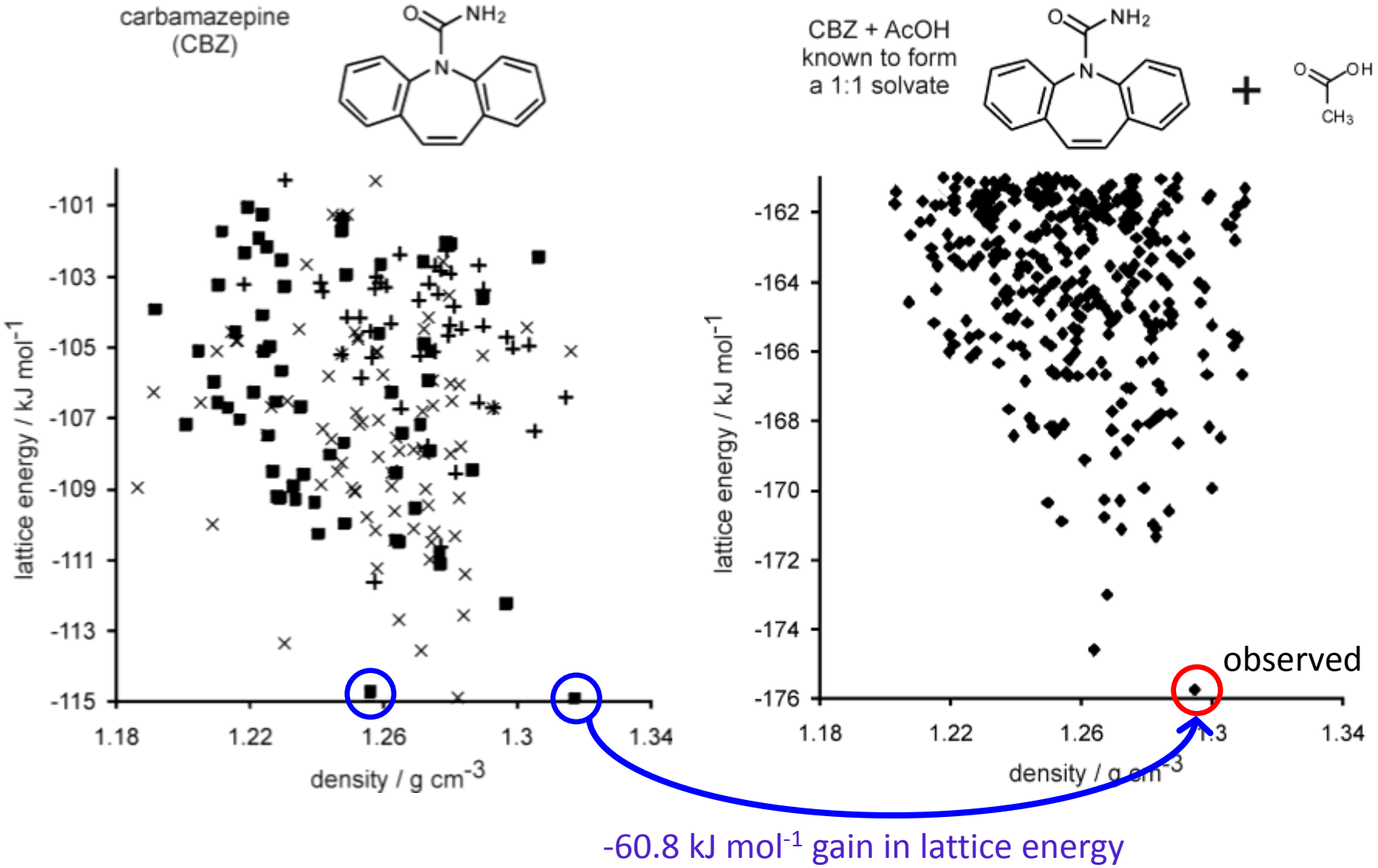
<u>Questions that we should ask of computational methods</u>:

- If we know that a co-crystal / solvate will form, and we know its composition (stoichiometry) can we predict its crystal structure?
- Could we have predicted the stoichiometry?
- Can we predict if a 2-component structure will form at all?

2-component structure of known composition

CBZ + AcOH
known to form
a 1:1 solvate





Compare to acetic acid in pure form:

$\Delta H°_{vap}$ = 51.6 ± 1.5 kJ mol^{-1}

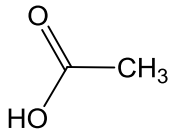
lattice energy (calc) = -58.1 kJ mol^{-1}

J. Amer Chem. Soc. (2006), 128, 14466-14467.

Methods carry over to 2 components. We will get back to computational expense.

More challenging: 2-component structures of unknown composition.

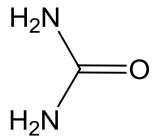
Acetic acid (AcOH)



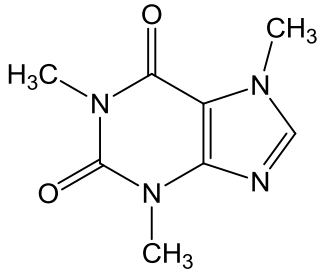
+

(M)

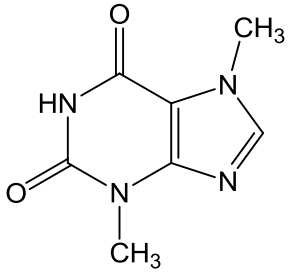
urea



caffeine



theobromine



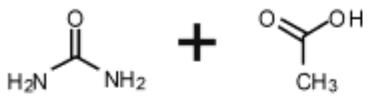
i) Predict all possible crystal structures at a range of stoichiometries

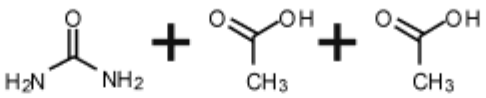
1:0 (neat crystal); 1:1; 1:2, etc. (M:AcOH)

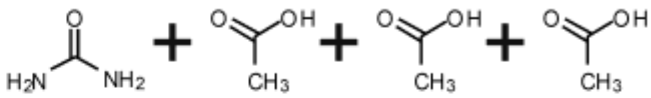
Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.

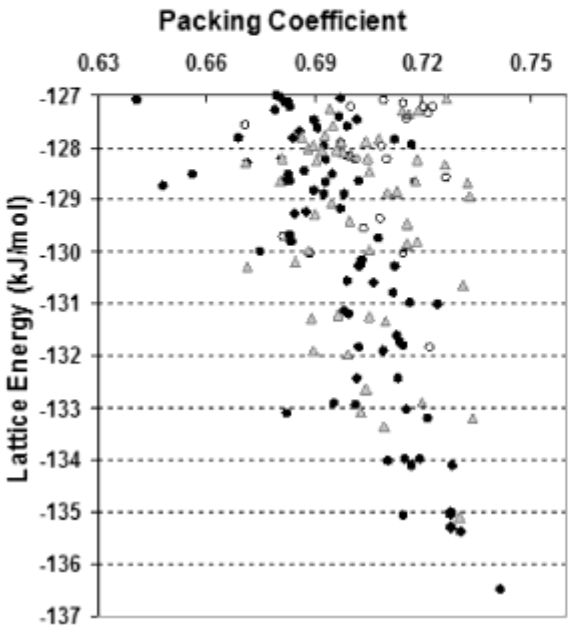
(urea:AcOH)

ratio 1 : 1 1 : 2 1 : 3

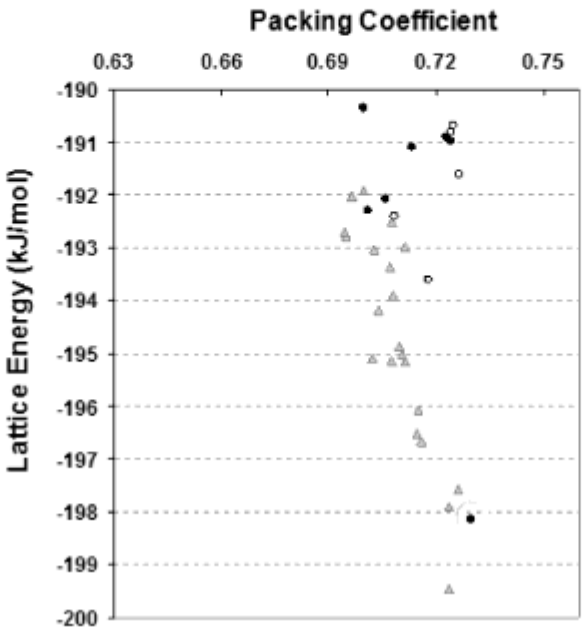




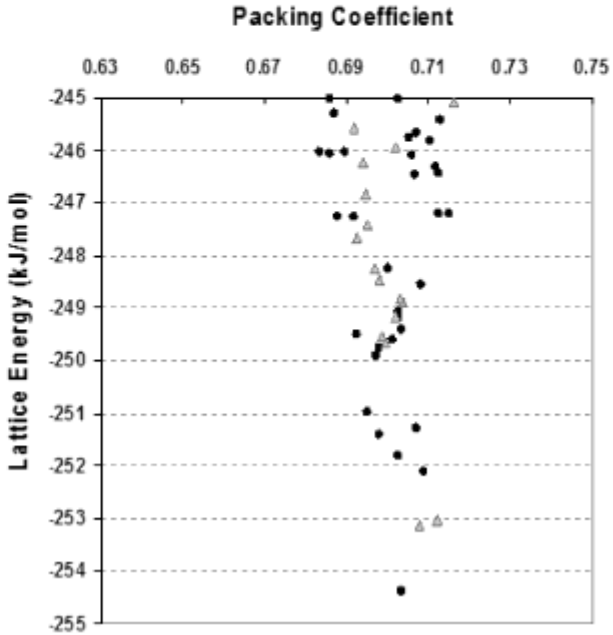




-136.5 kJ mol^{-1}



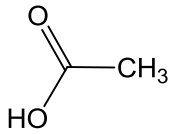
-199.5 kJ mol^{-1}



-254.3 kJ mol^{-1}

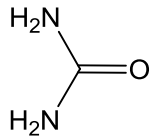
Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.

Acetic acid (AcOH)

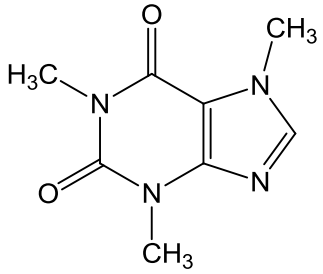


+

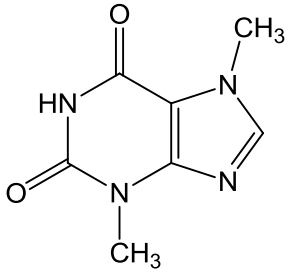
urea



caffeine



theobromine



(M)

i) Predict all possible crystal structures at a range of stoichiometries

1:0 (neat crystal); 1:1; 1:2, etc. (M:AcOH)

ii) Assess relative stability at <u>constant composition</u>

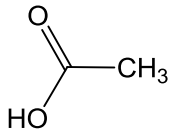
no co-crystallisation: $1 \cdot E^{M}_{latt,global\ min} + 2 \cdot E^{AcOH}_{latt}$

1:1 co-crystallisation: $1 \cdot E^{M:AcOH}_{latt,global\ min} + 1 \cdot E^{AcOH}_{latt}$

1:2 co-crystallisation: $1 \cdot E^{M:2AcOH}_{latt,global\ min} + 0 \cdot E^{AcOH}_{latt}$

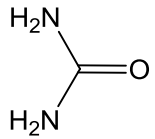
Chem. Commun. (2010), <u>46</u>, 2224-2226; *Chem. Eur. J.* (2008), <u>14</u>, 8830-8836.

Acetic acid (AcOH)

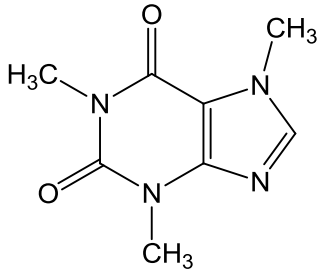


+

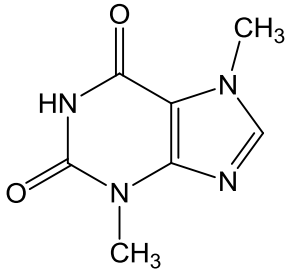
urea



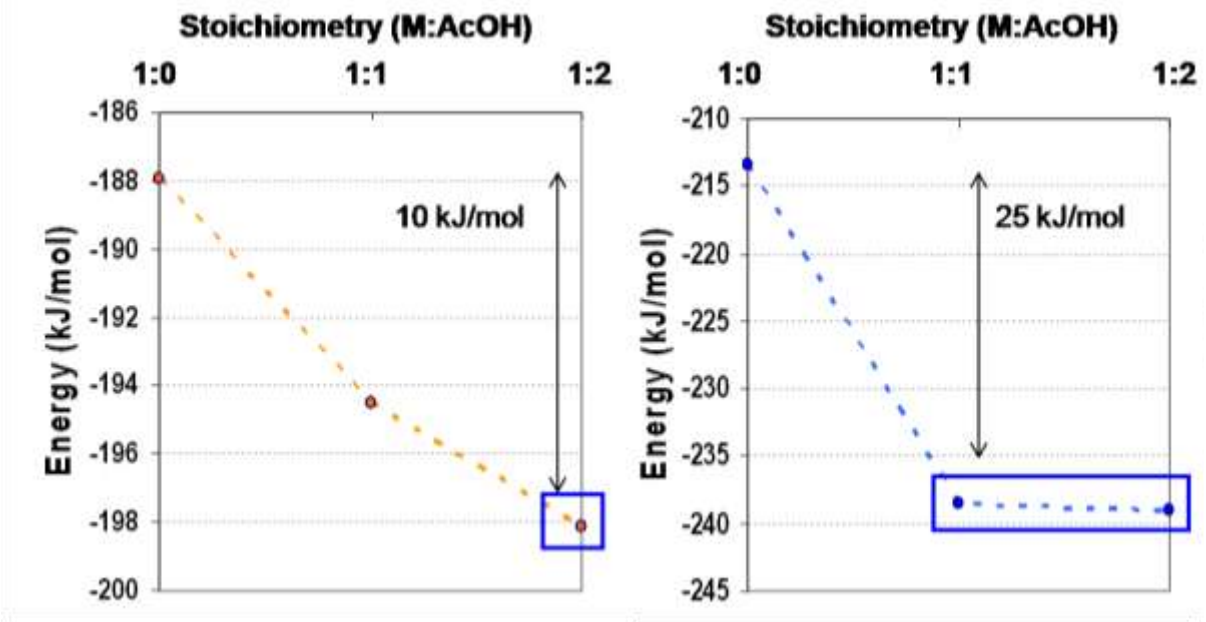
caffeine



theobromine



(M)



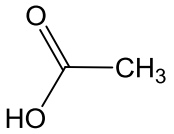
Observed as 1:2

Two forms known: 1:1 and 1:2

Material from liquid assisted grinding (LAG): cannot grow single crystal

Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.

Acetic acid (AcOH)



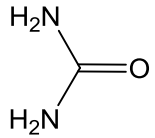
+

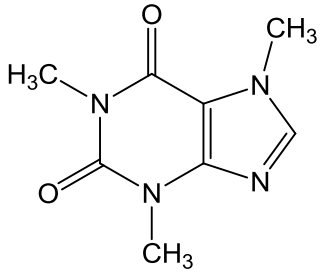
urea

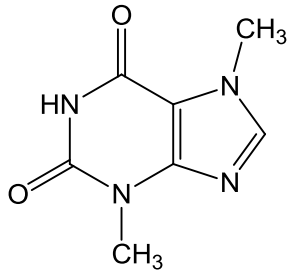
caffeine

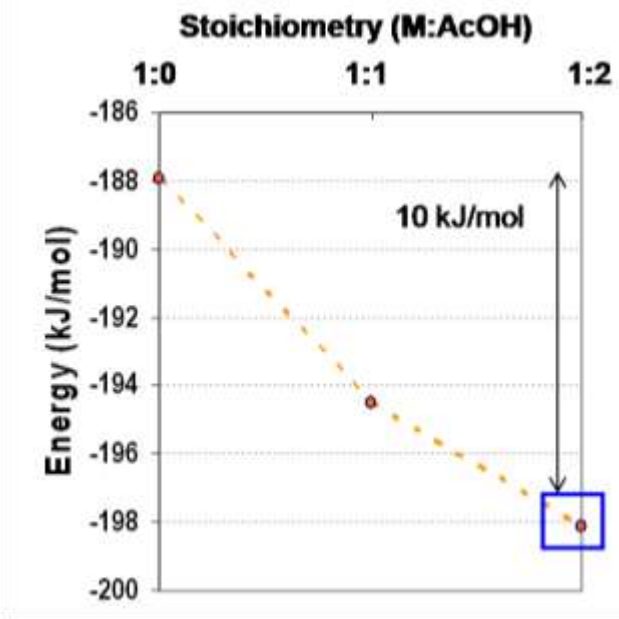
theobromine

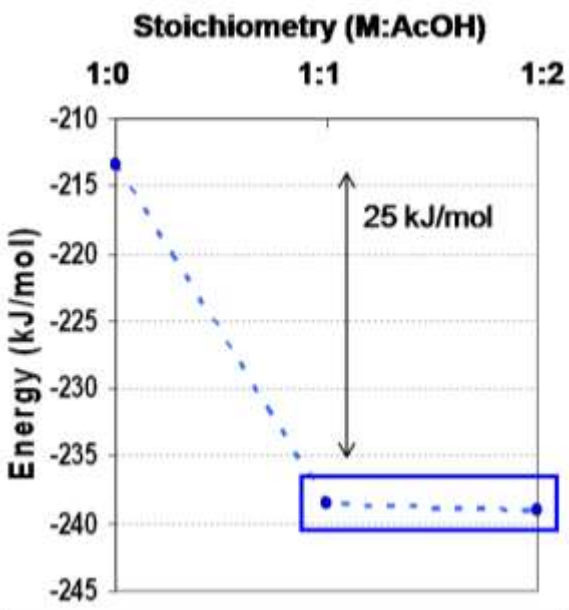
(M)

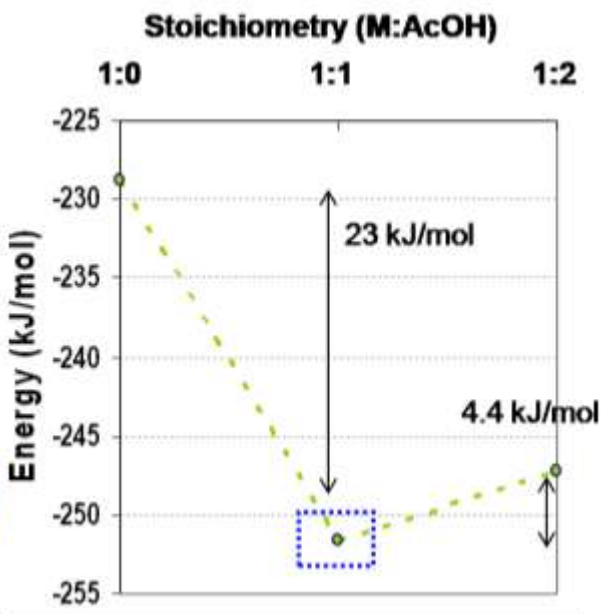










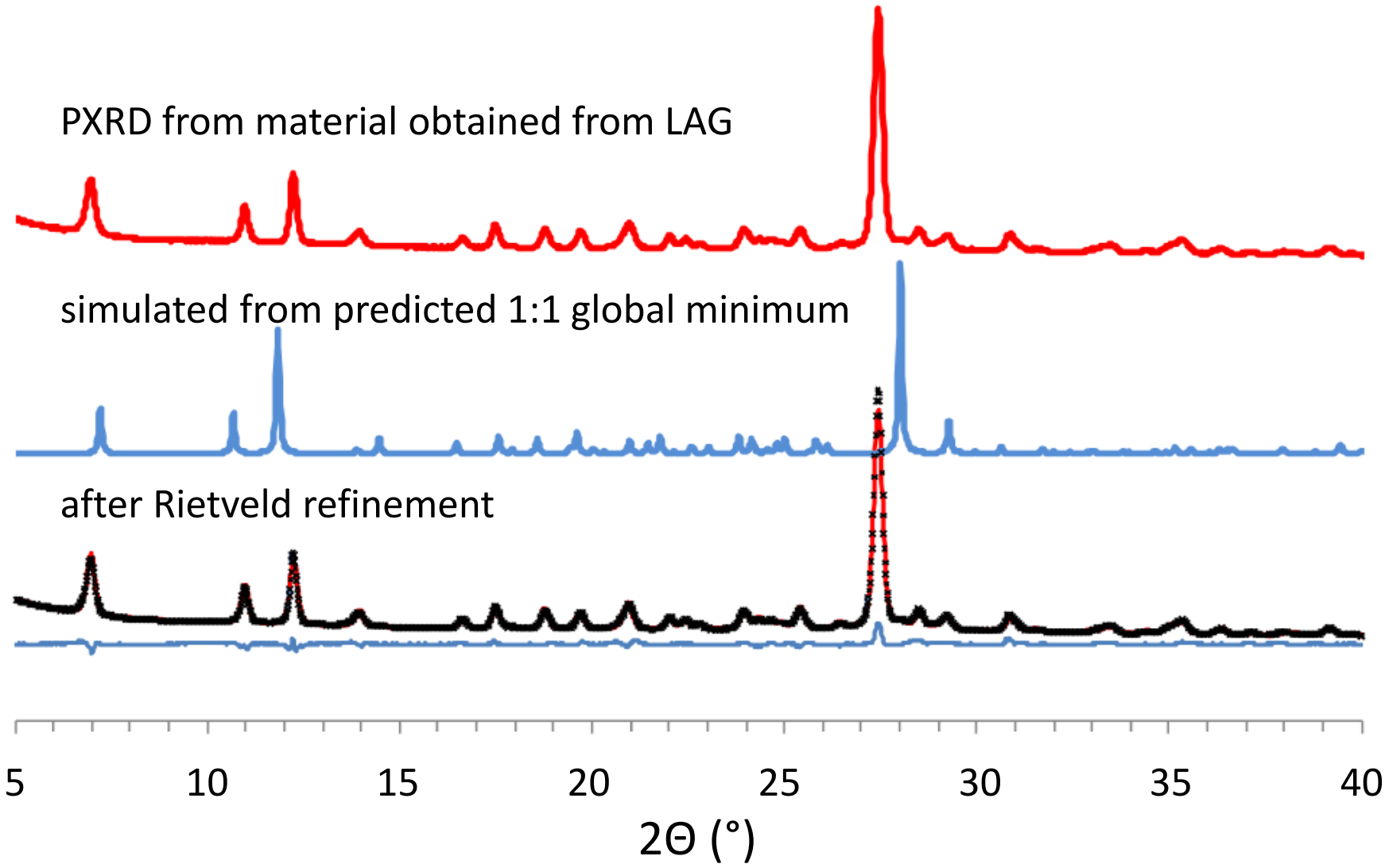


Observed as 1:2

Two forms known: 1:1 and 1:2

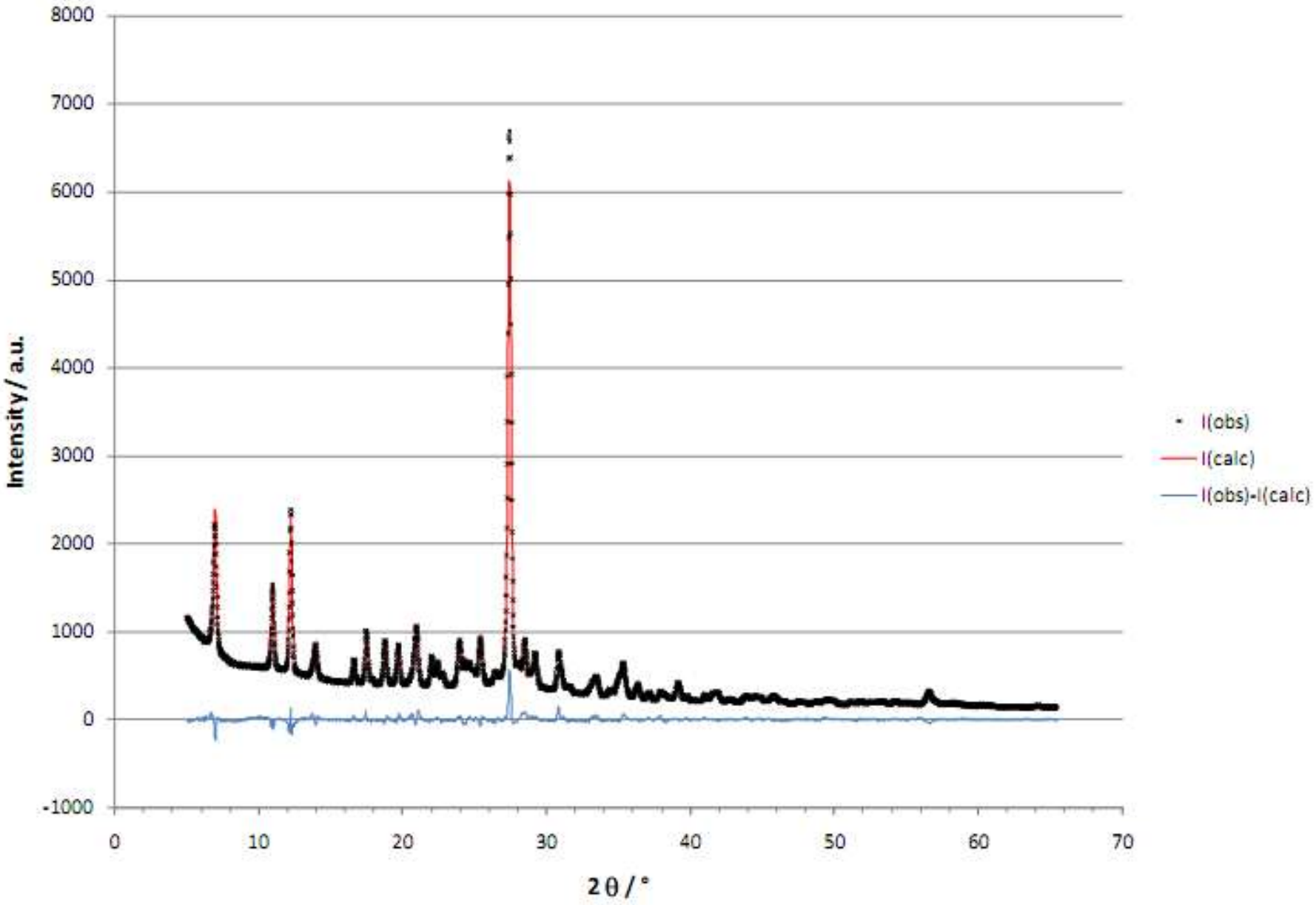
Predicted to form 1:1

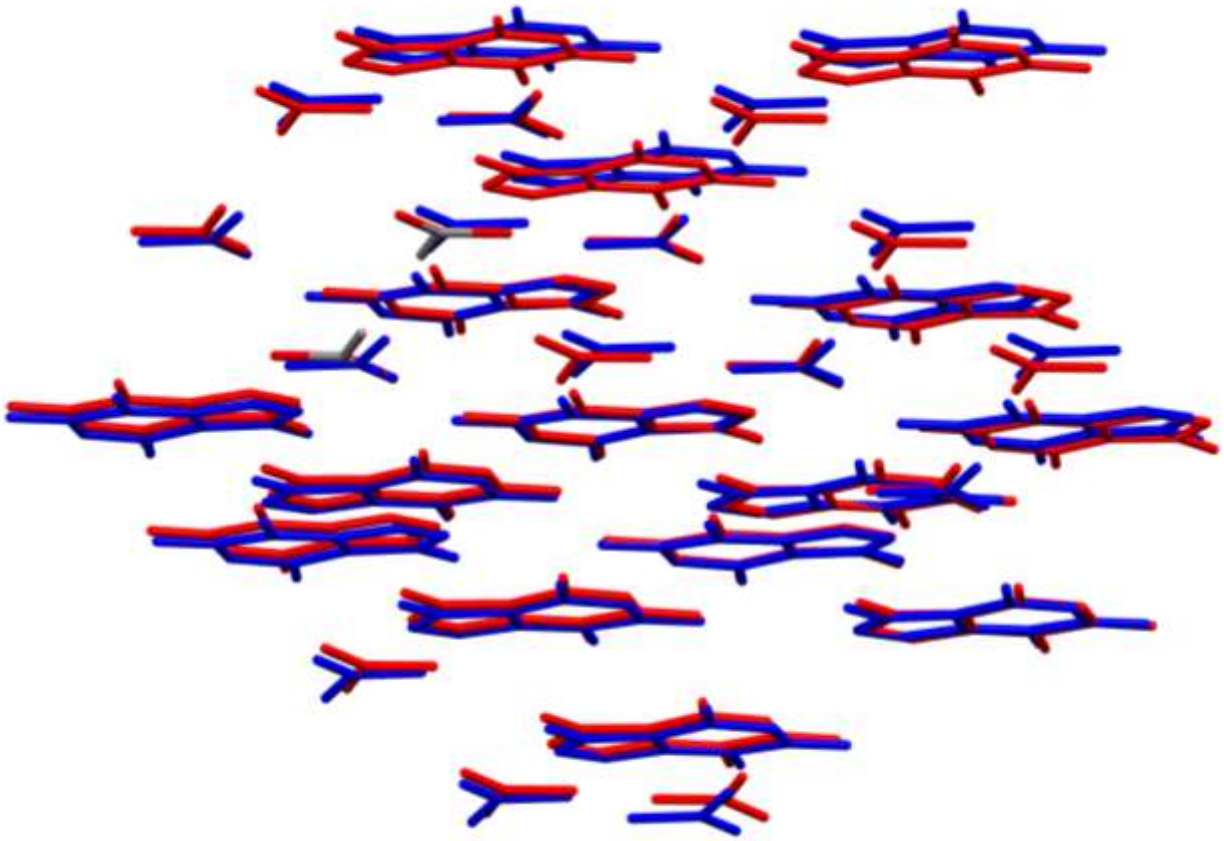
Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.



Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.

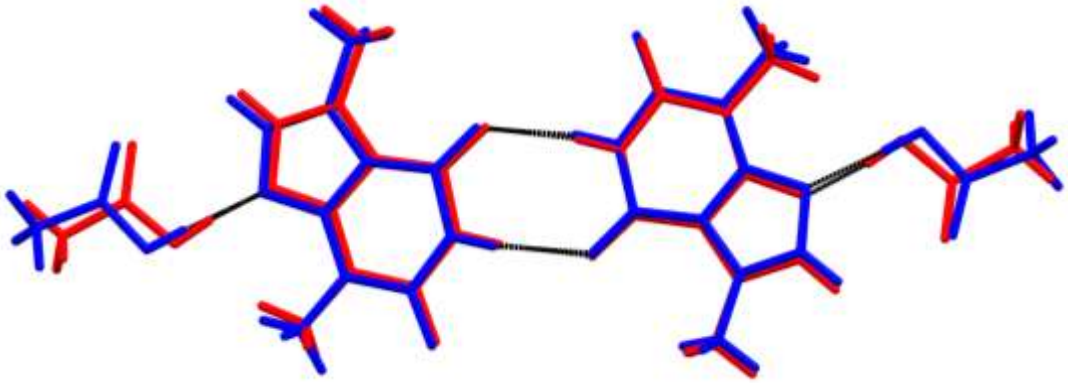
1:1 theobromine : acetic acid Rietveld refinement





Red = global minimum predicted

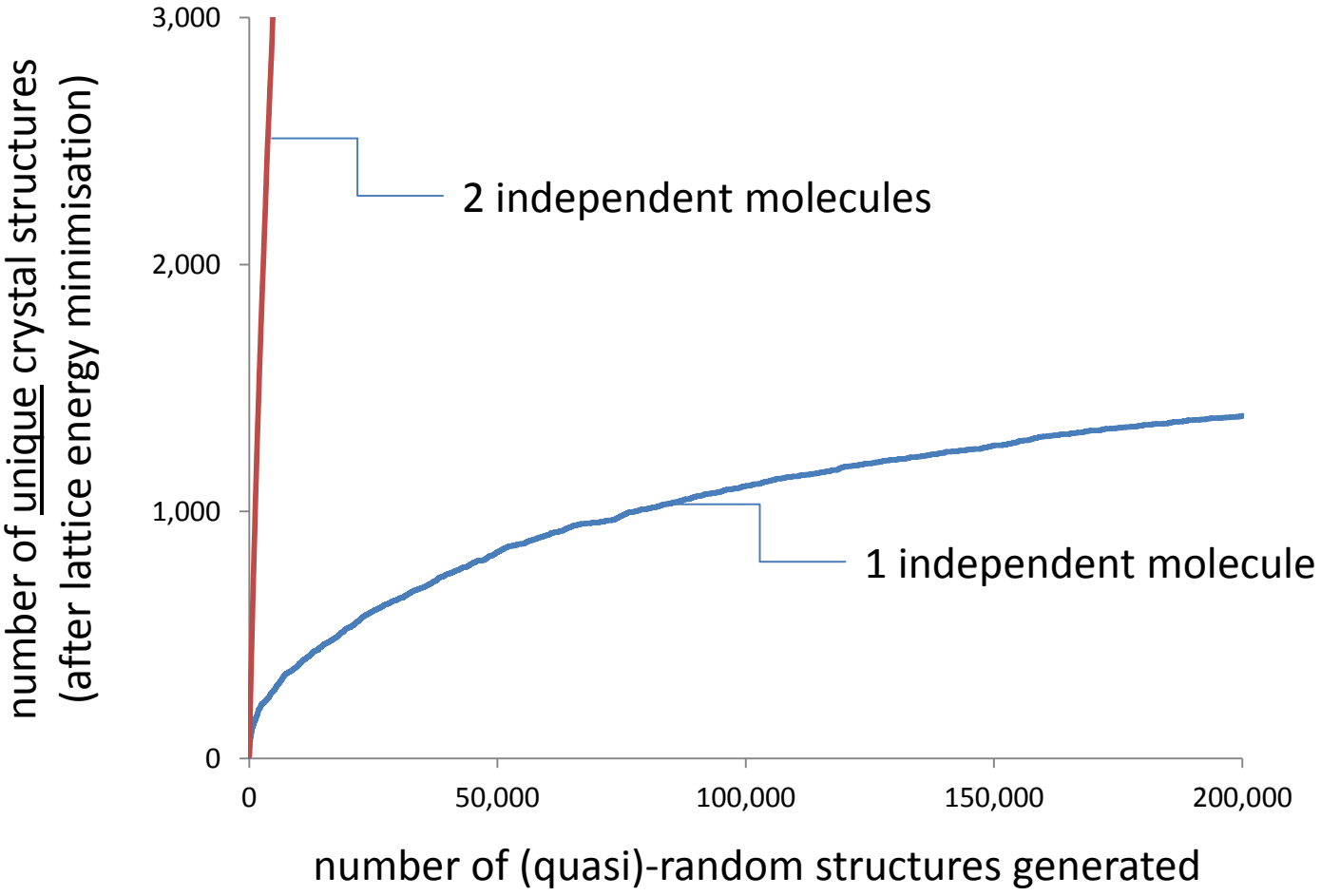
Blue = refinement to PXRD



Chem. Commun. (2010), 46, 2224-2226; *Chem. Eur. J.* (2008), 14, 8830-8836.

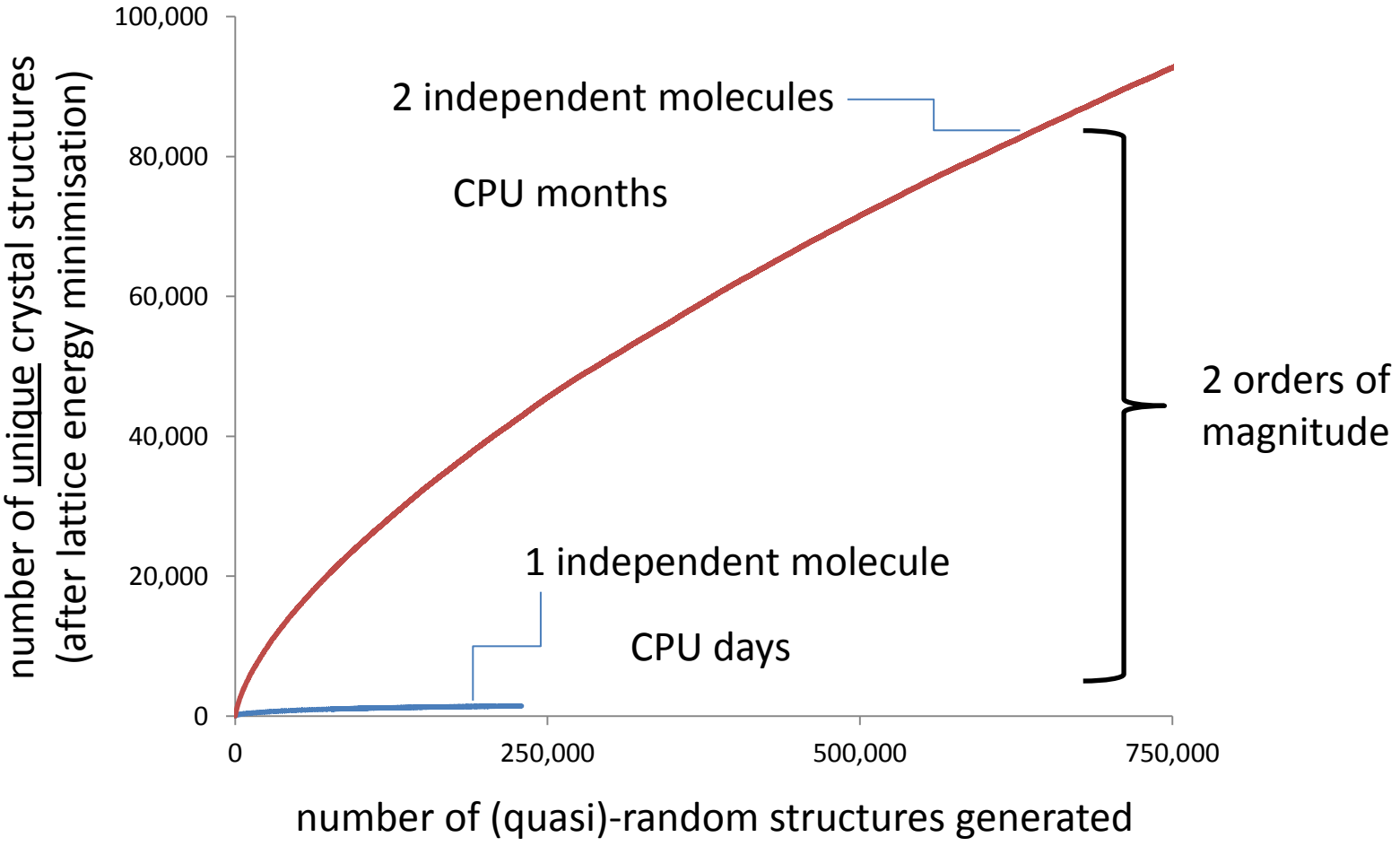
This approach seems to work (surprisingly) well…

… but the calculations involved are expensive!



This approach seems to work (surprisingly) well…

… but the calculations involved are expensive!



What can we say about co-crystal or solvate formation?

Our questions:

- If we know that a solvate will form, and we know its composition (main molecule : solvent stoichiometry) can we predict its crystal structure?

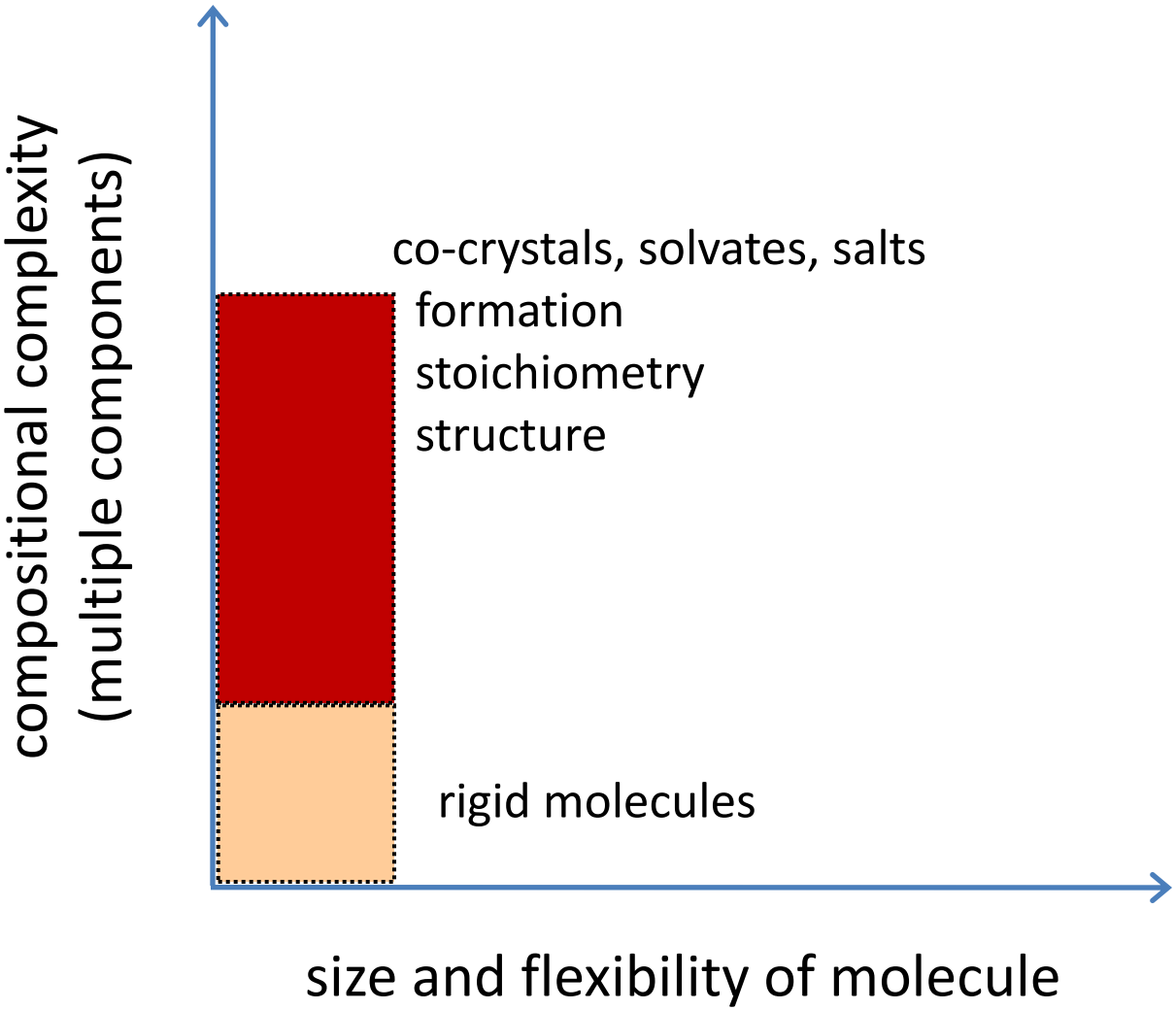
- Could we have predicted the stoichiometry?

- Can we predict if a solvate will form at all?

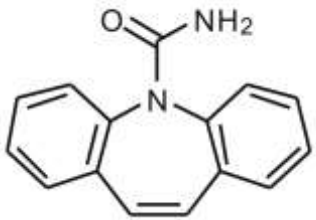
Promising results so far.

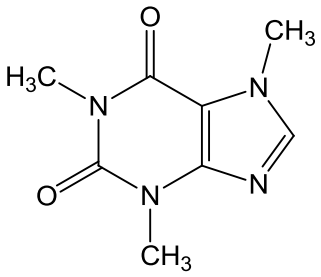
Energy differences are very small. (Entropy has been largely ignored so far.)

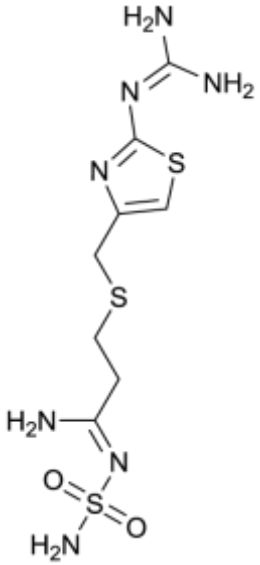
The calculations are expensive!

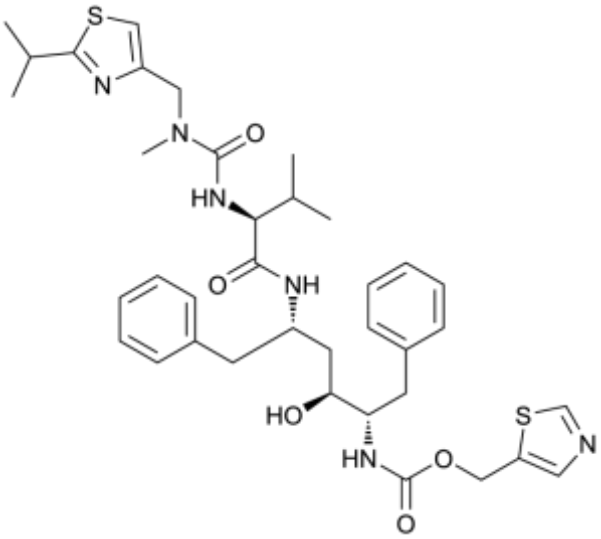

compositional complexity
(multiple components)
co-crystals, solvates, salts
formation
stoichiometry
structure
rigid molecules
size and flexibility of molecule

Dealing with molecular flexibility



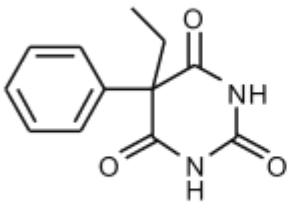




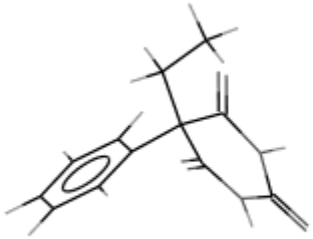


Crystal Structure Prediction
Flexible Molecules

molecular connectivity



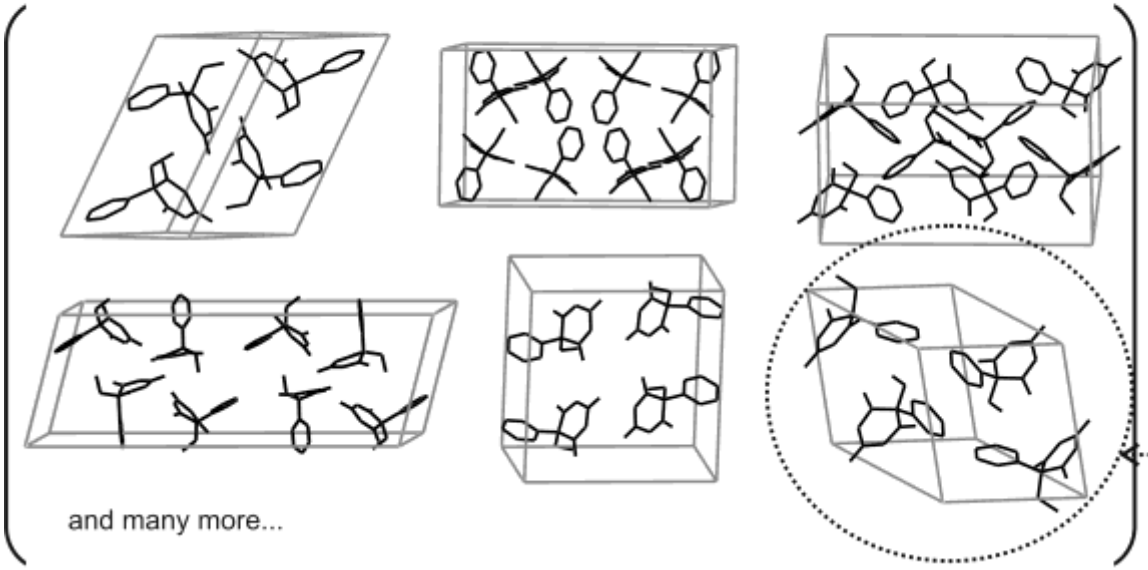
3D molecular structure



Now we need to consider all possible conformations

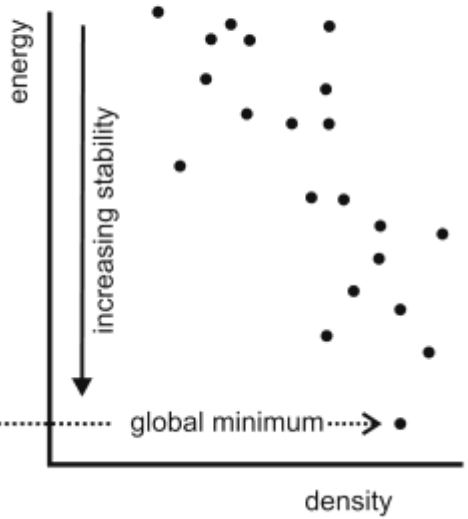
crystal structure search

packing possibilities

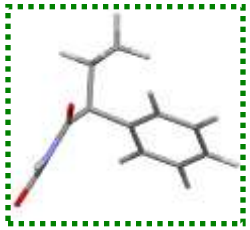


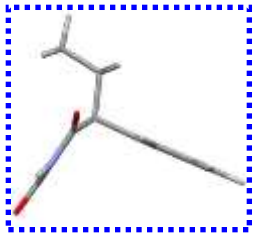
energy and property calculations

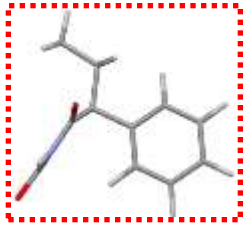
ranking of structures

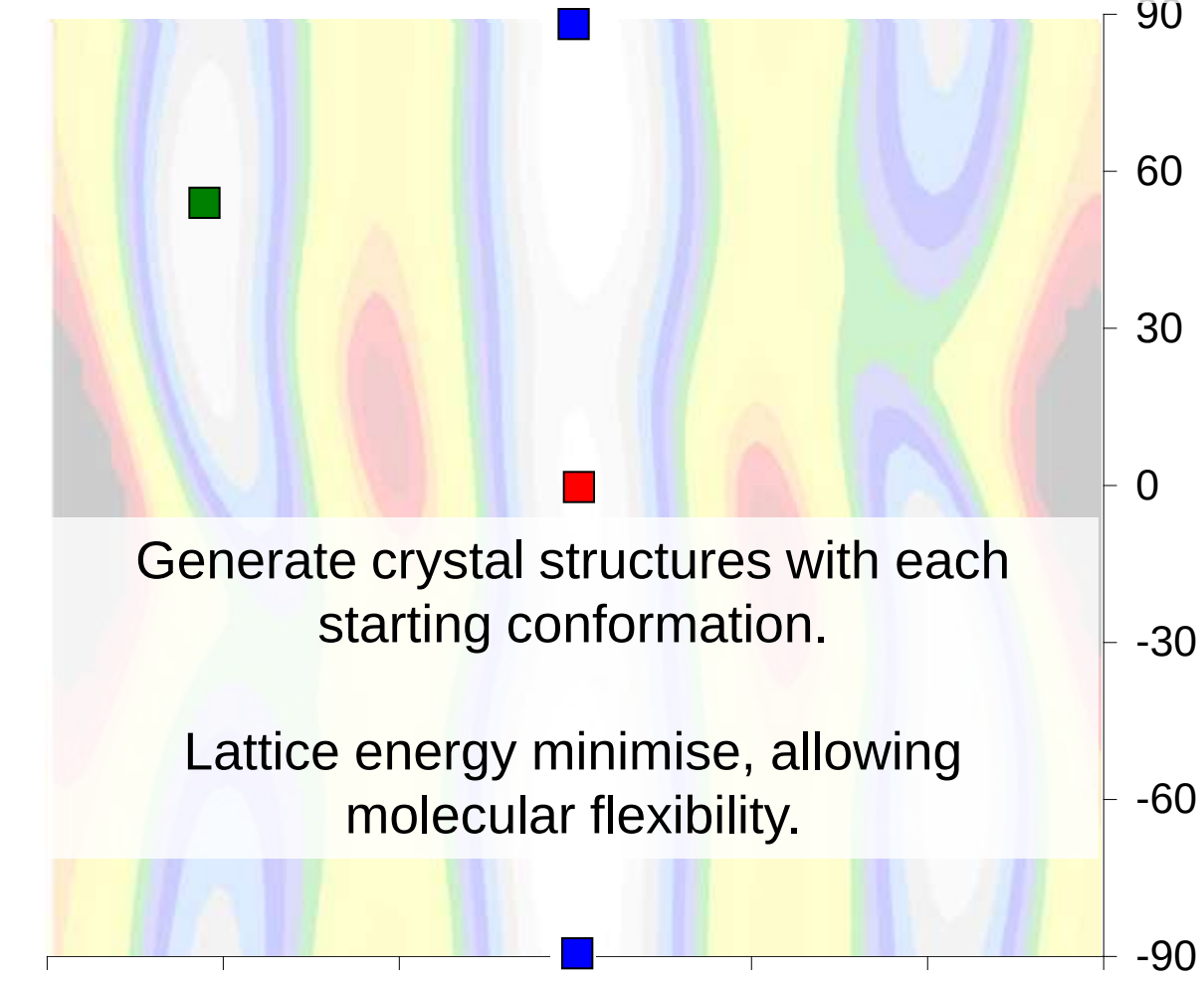


We can do this exhaustively for a few degrees of flexibility







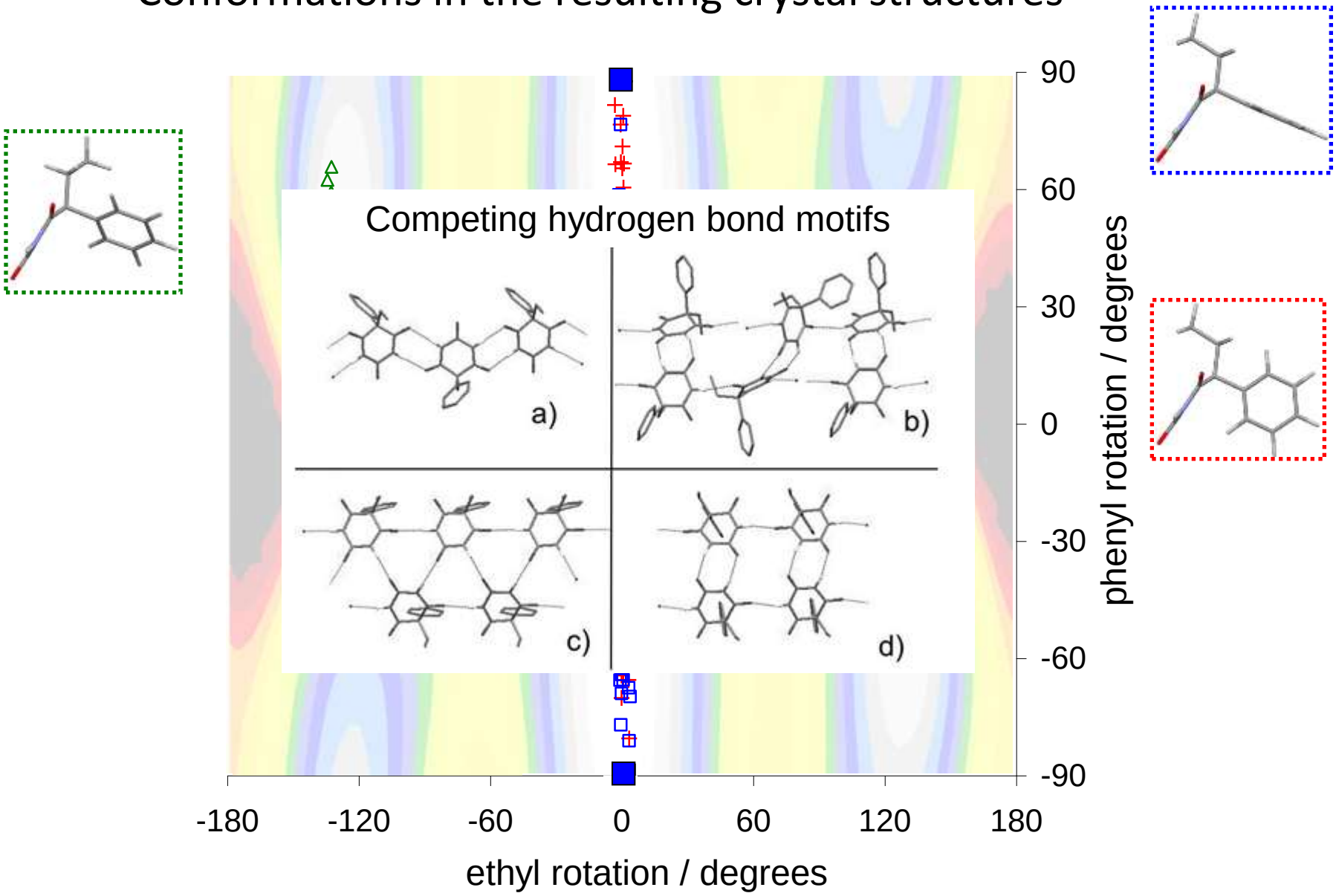


Generate crystal structures with each starting conformation.

Lattice energy minimise, allowing molecular flexibility.

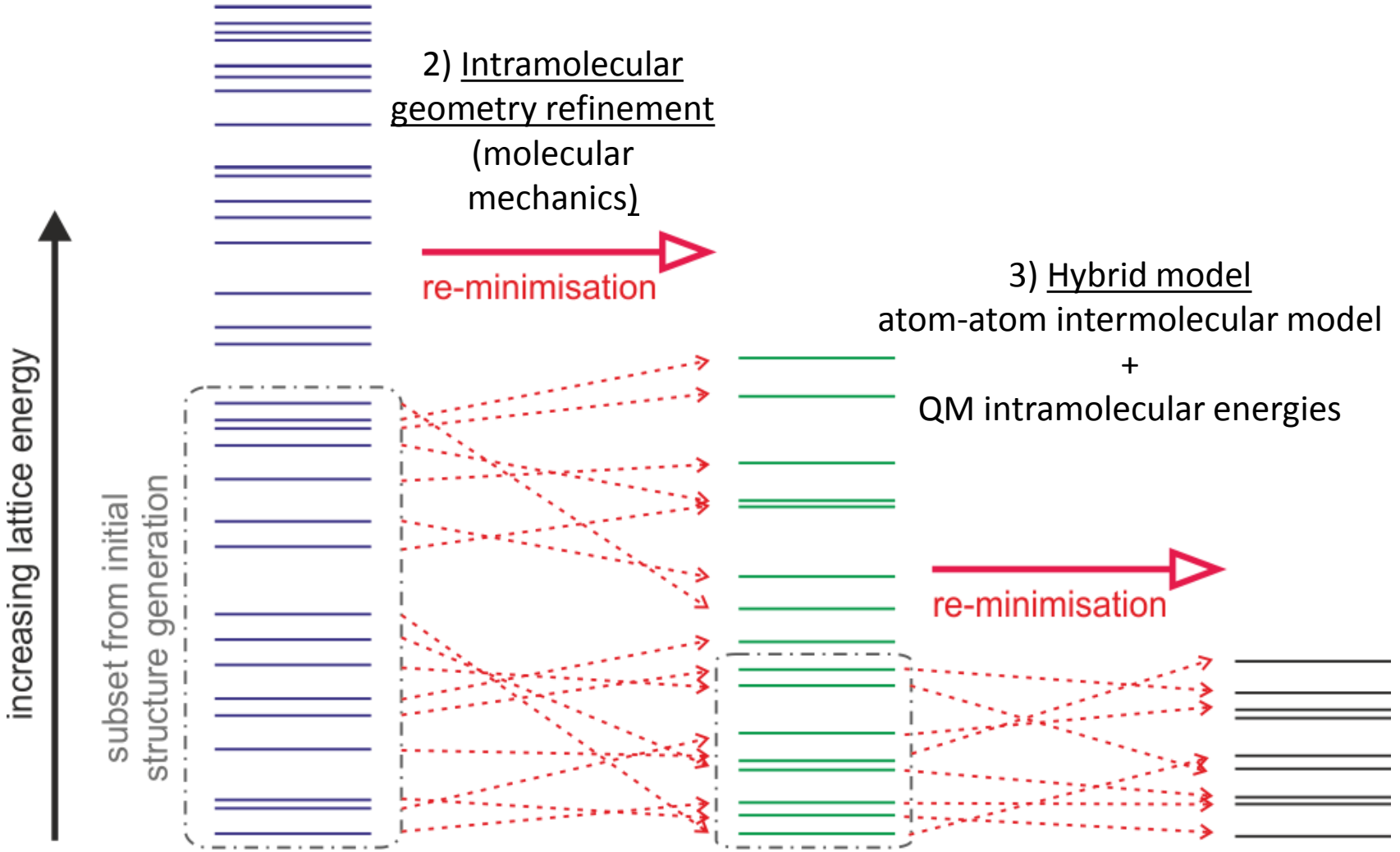
contour lines each represent 4 kJ mol^{-1}

Conformations in the resulting crystal structures



contour lines each represent 4 kJ mol^{-1}

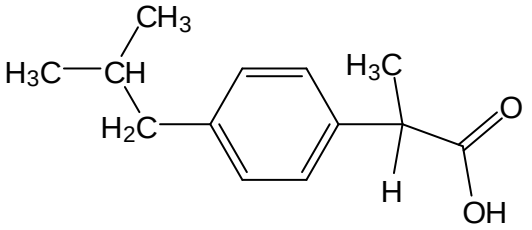
Hierarchical approach for flexible molecules

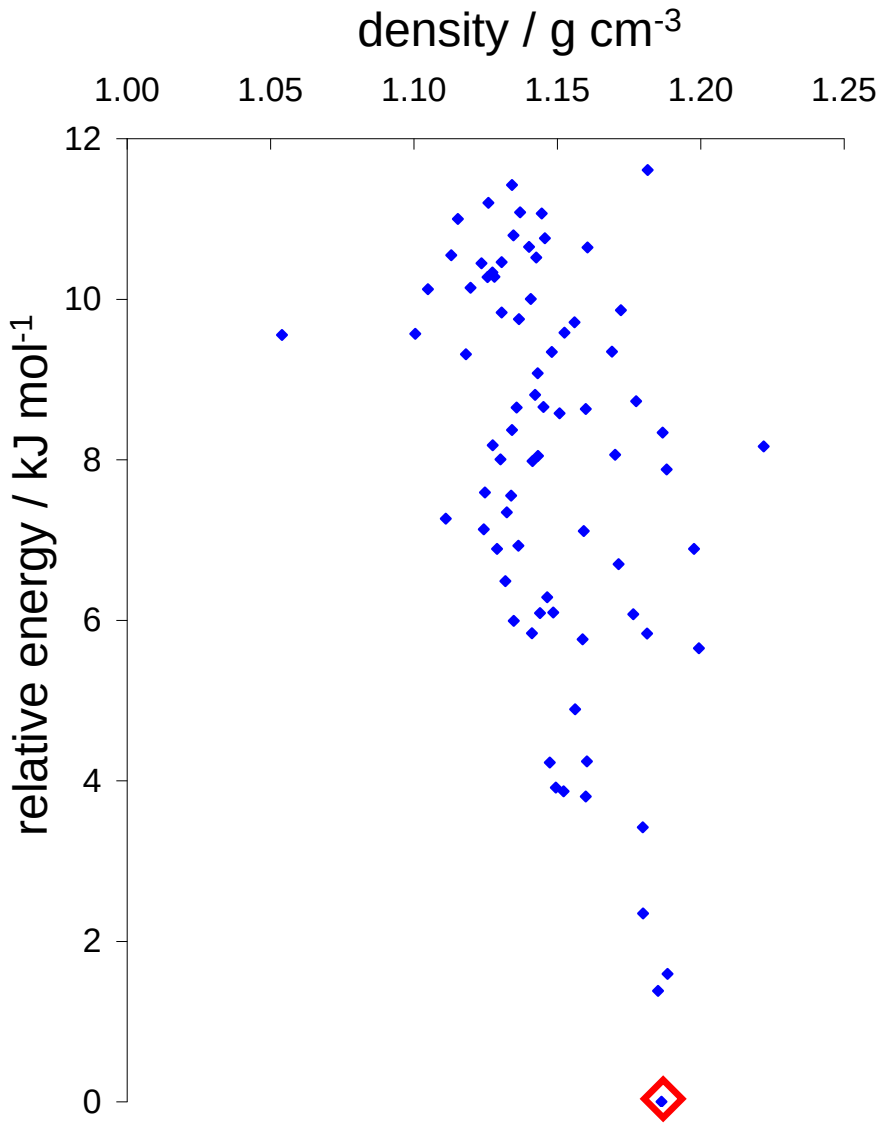


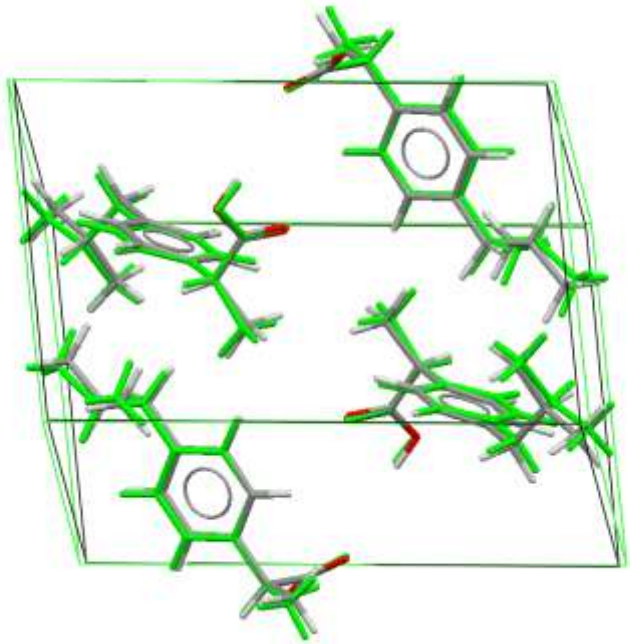
1) Crystal structure search
Rigid conformations
simple intermolecular potential

$$U_{total} = U_{molecular}(QM, \rho) + \sum_{i \in M, k \in N} \left[A_{ik} \exp(-B_{ik} R_{ik}) - C_{ik} R_{ik}^{-6} + U_{electr}(\rho_M, \rho_N) \right]$$

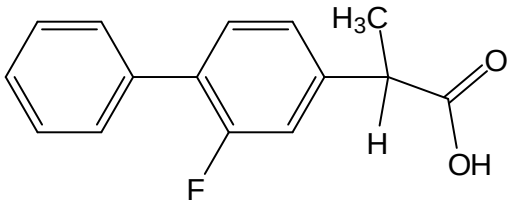
Ibuprofen

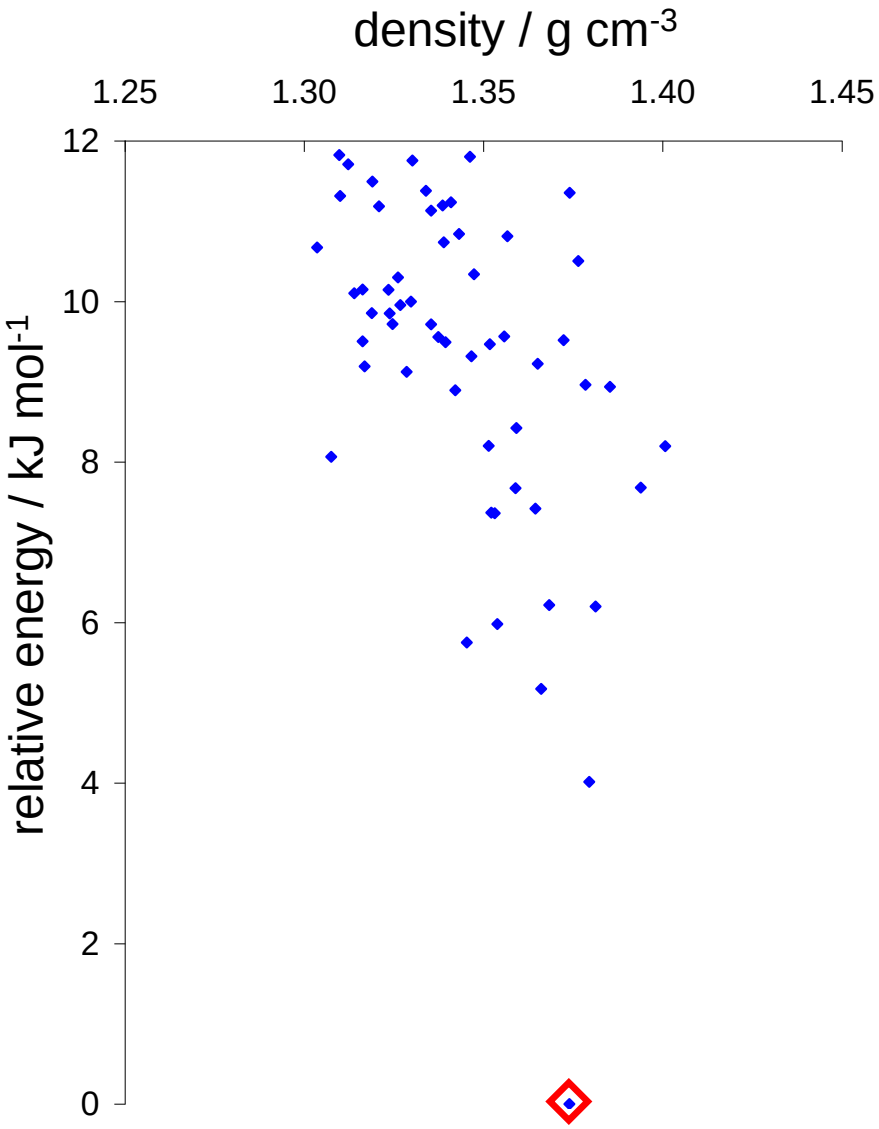

CH3
H3C
CH
H2C
H3C
O
H
OH

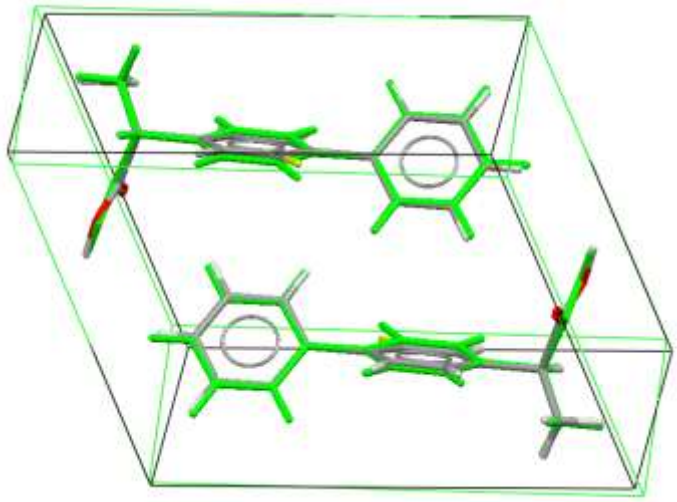

density / g cm-3
1.00
1.05
1.10
1.15
1.20
1.25
relative energy / kJ mol-1
12
10
8
6
4
2
0



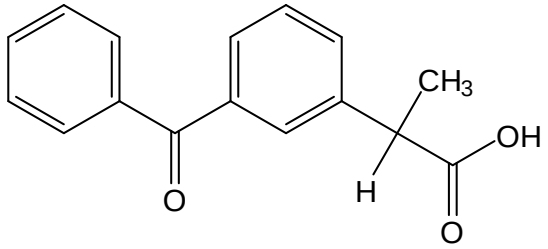
Flurbiprofen

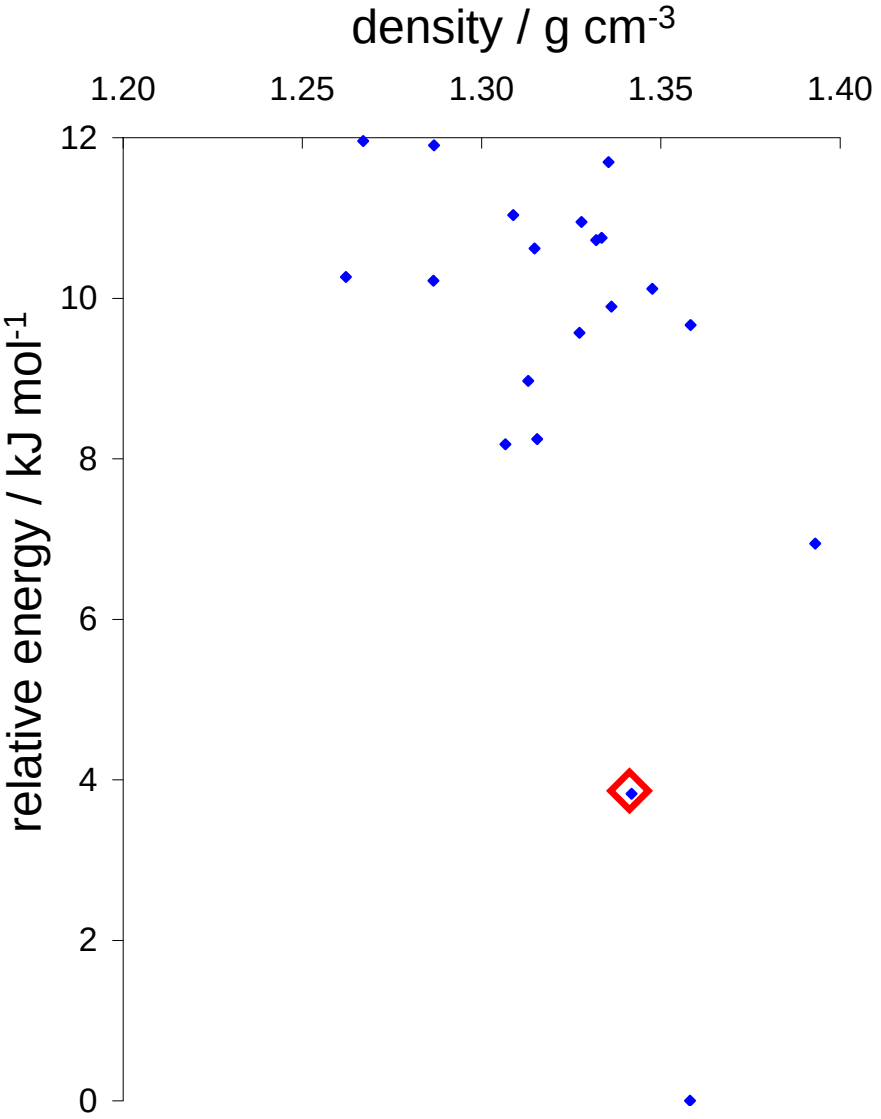


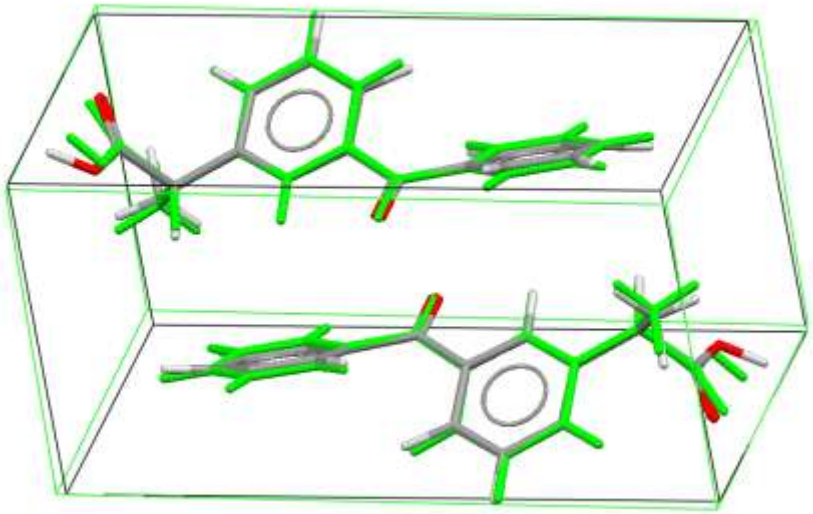




Ketoprofen

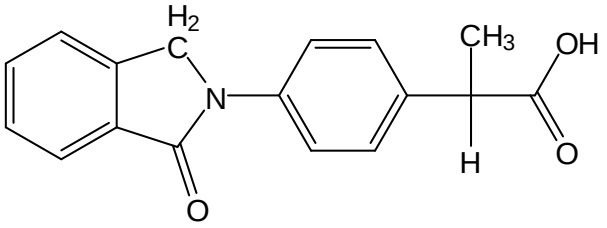


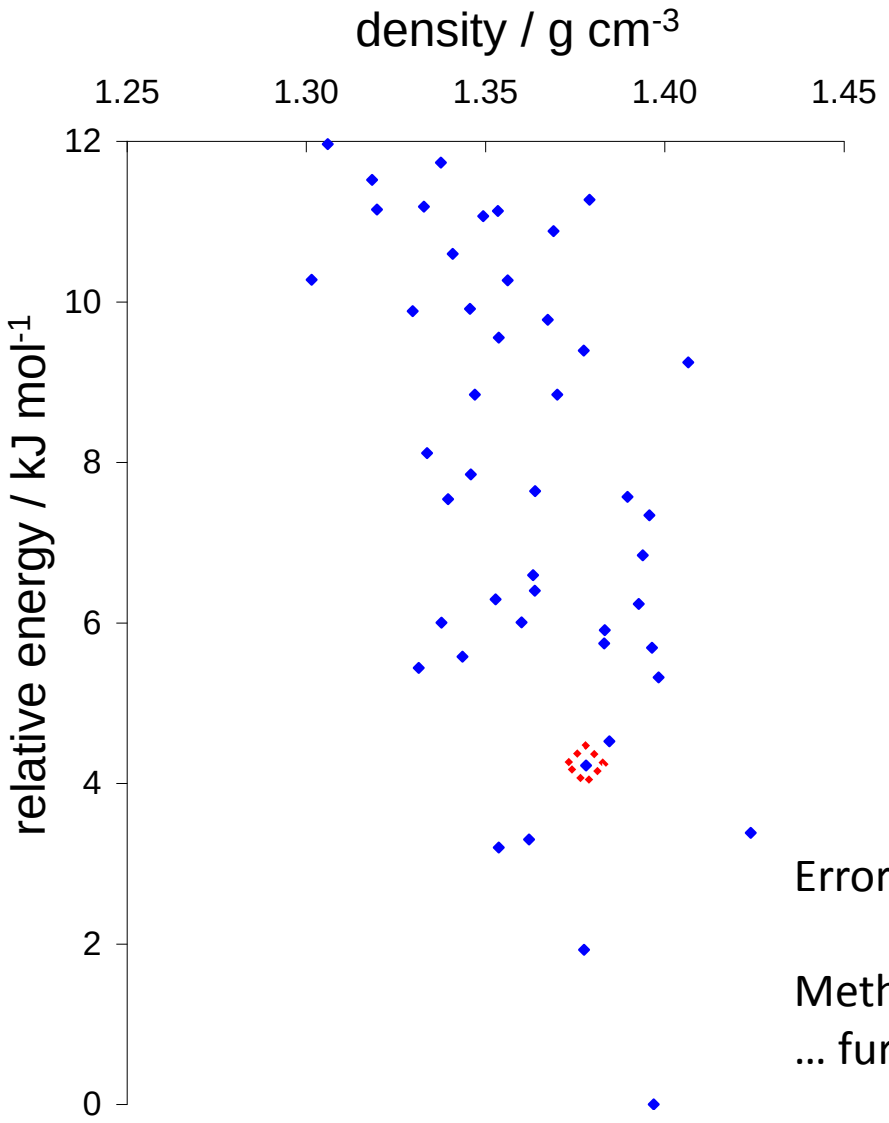


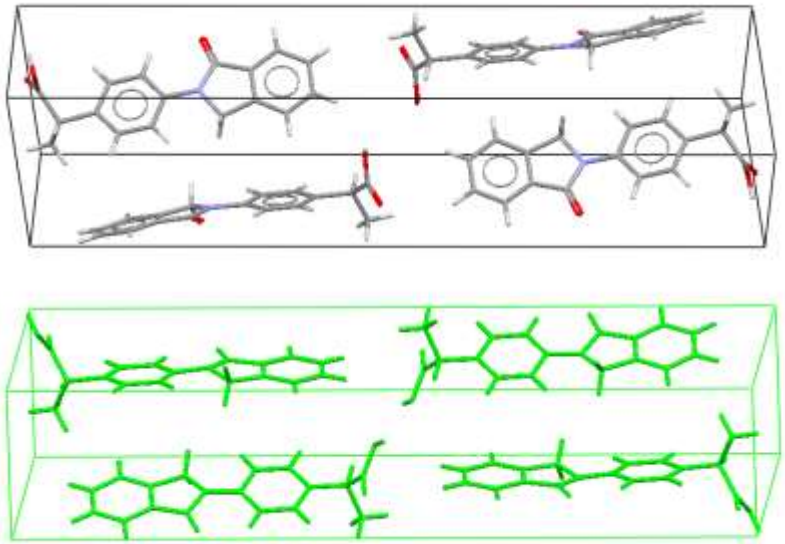


Polymorphism?
Imperfect energy model?

Indoprofen





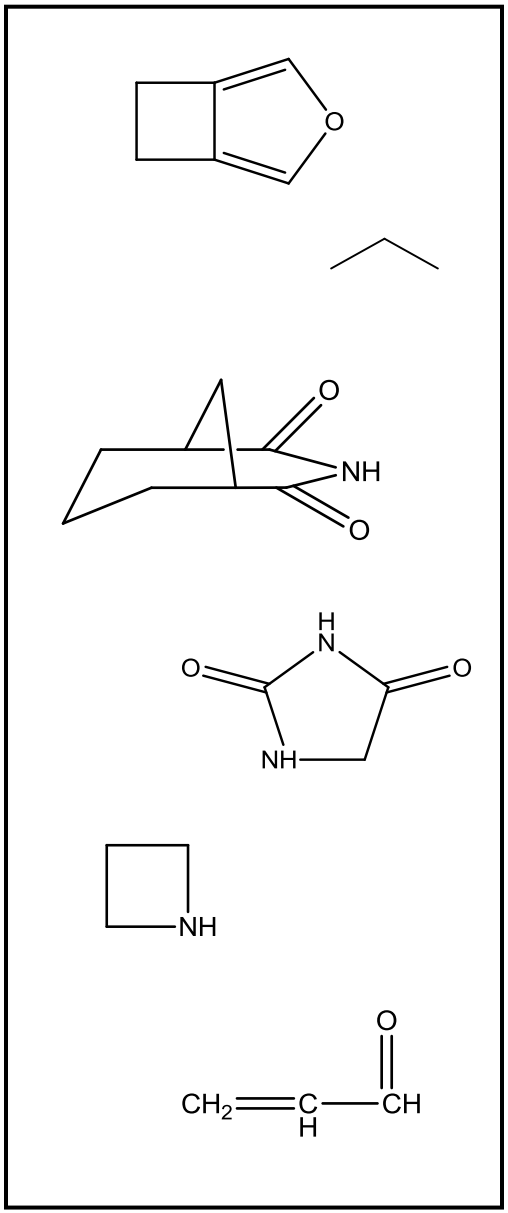


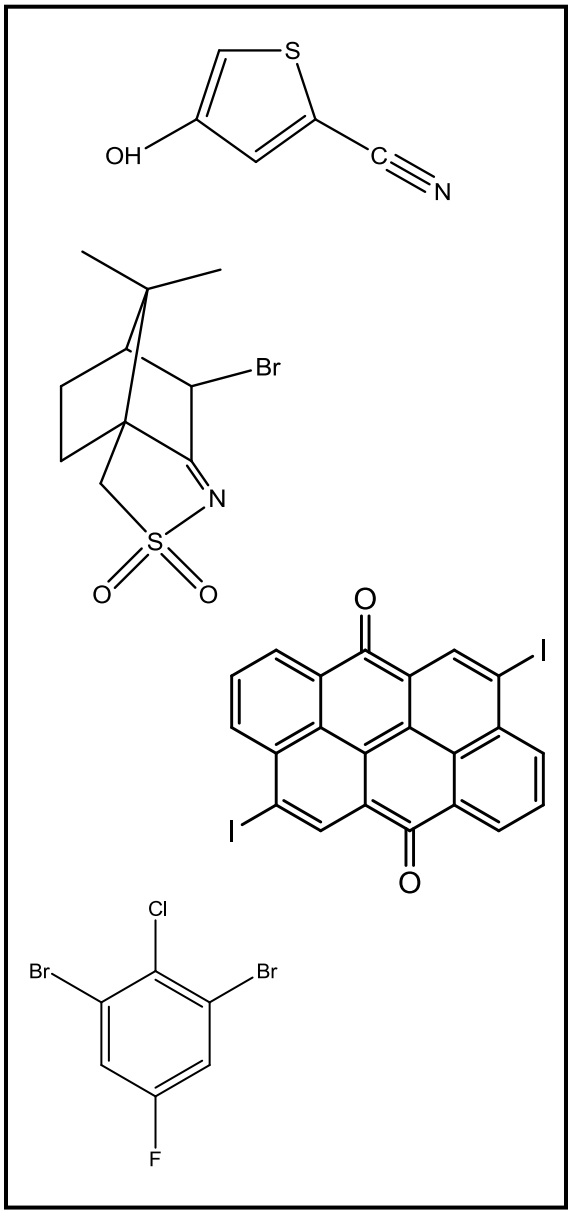
Error in molecular geometry

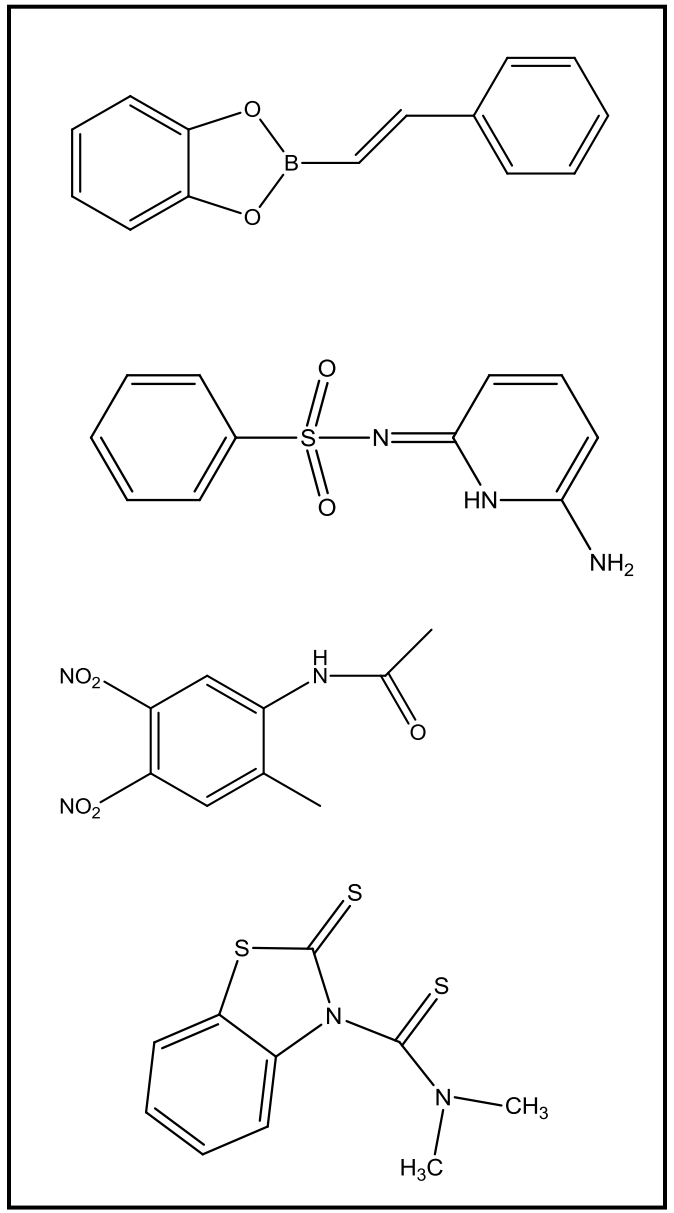
Methods are not complete
… further development of optimisation methods

Karamertzanis, Kazantsev, Adjiman, Pantelides, Price
(Imperial College London, University College London)

“Blind Tests” of Crystal Structure Prediction







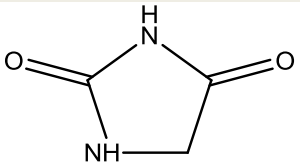
“Blind Tests” of Crystal Structure Prediction

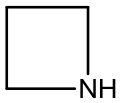
CSP1999: Lommerse *et al*, *Acta Cryst.* (2000), B56, 697.
CSP2001: Motherwell *et al*, *Acta Cryst.* (2002), B58, 647.
CSP2004: Day *et al*, *Acta Cryst.* (2005), B61, 511.
CSP2007: Day *et al*, Acta Cryst. (2009), B65, 107-125.

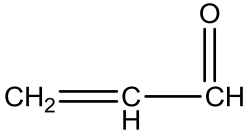
Little success in first 3 rounds.

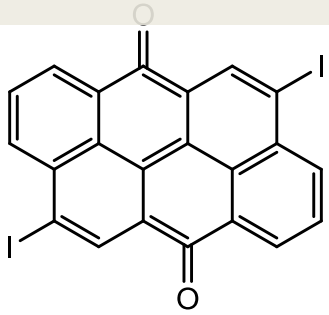
Increased success in 2007.
Several groups with successful predictions.

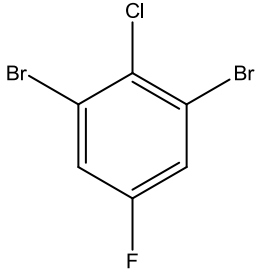
Molecules are all fairly rigid.

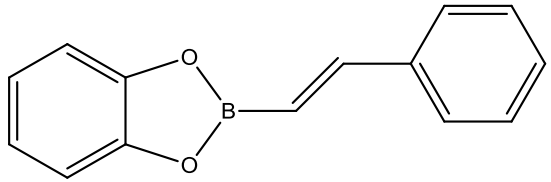


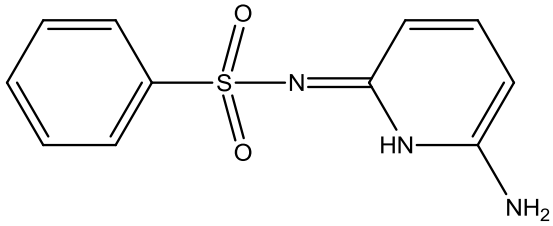


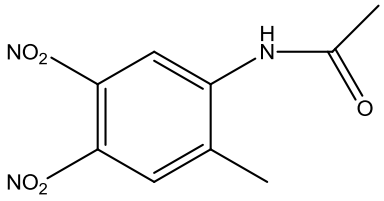


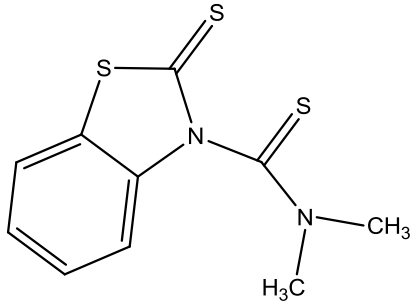








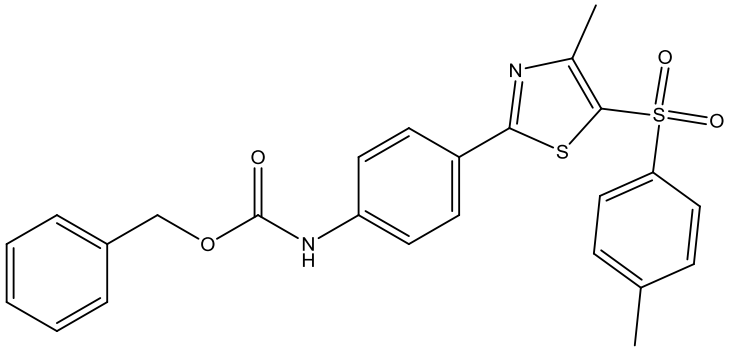




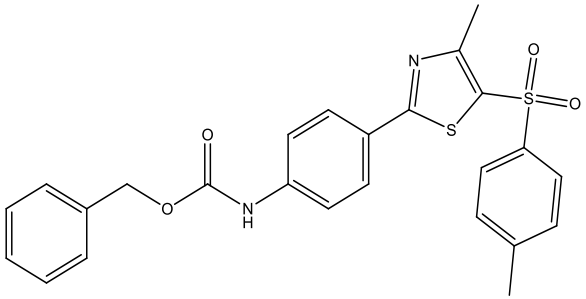
Latest 2010 blind test

Similar results to 2007 on small molecules. Publication in preparation.

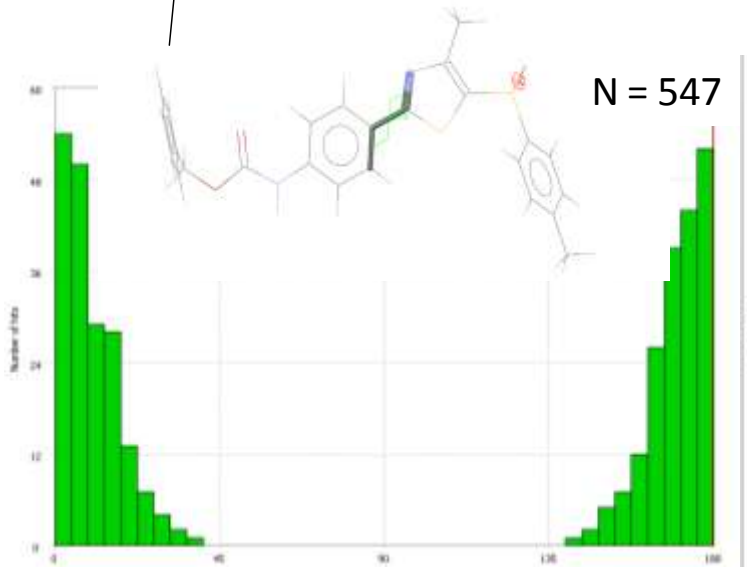
CSP2010 also included more challenging targets:

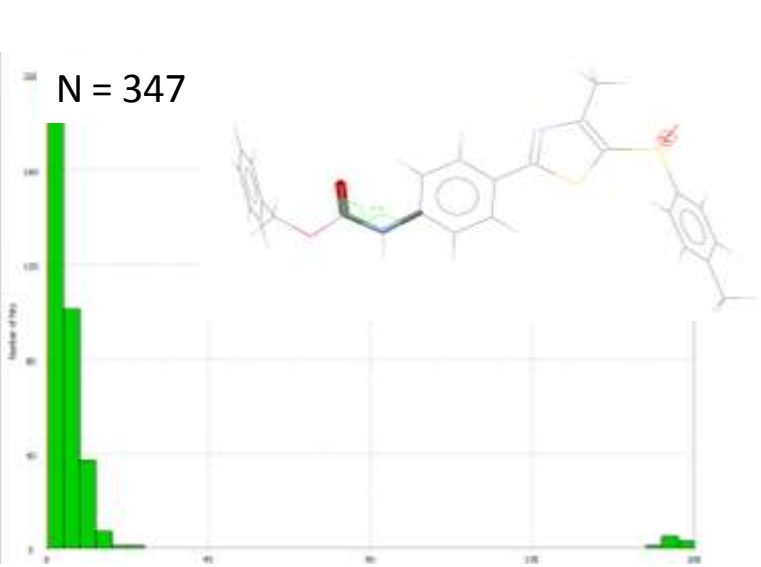


Int. J. Pharmaceutics (2011), in press, doi:10.1016/j.ijpharm.2011.03.058.

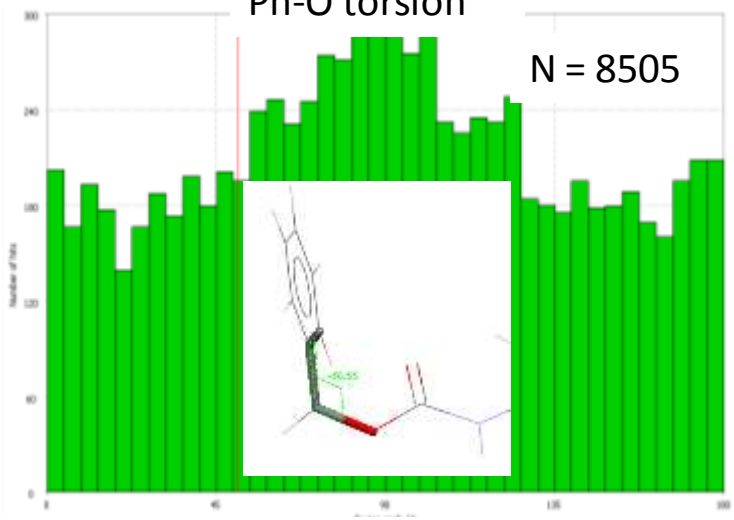


Exploring conformational space: Database guidance

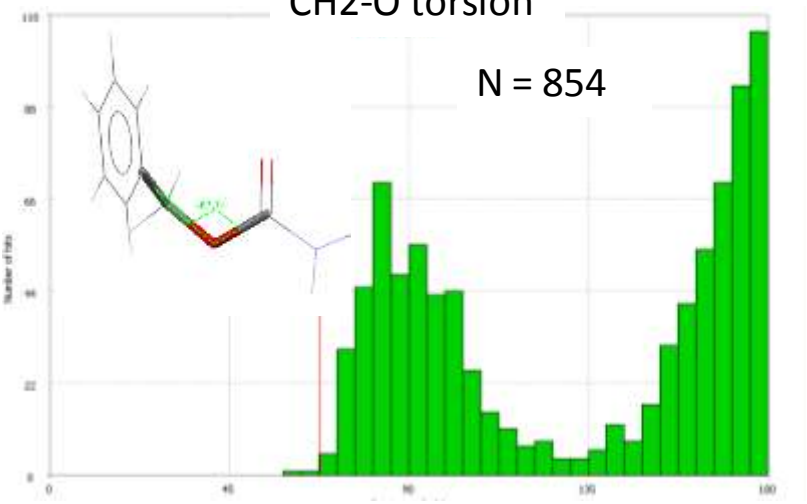




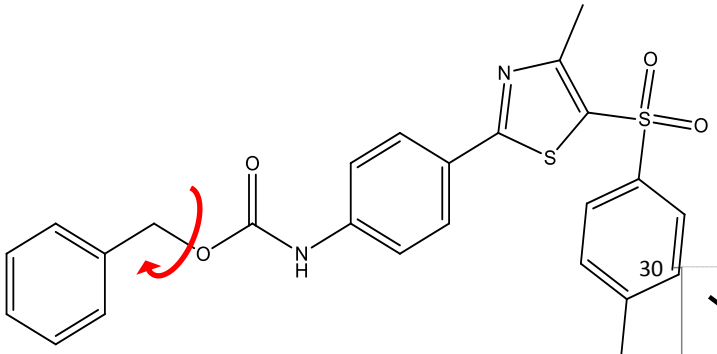
Ph-O torsion

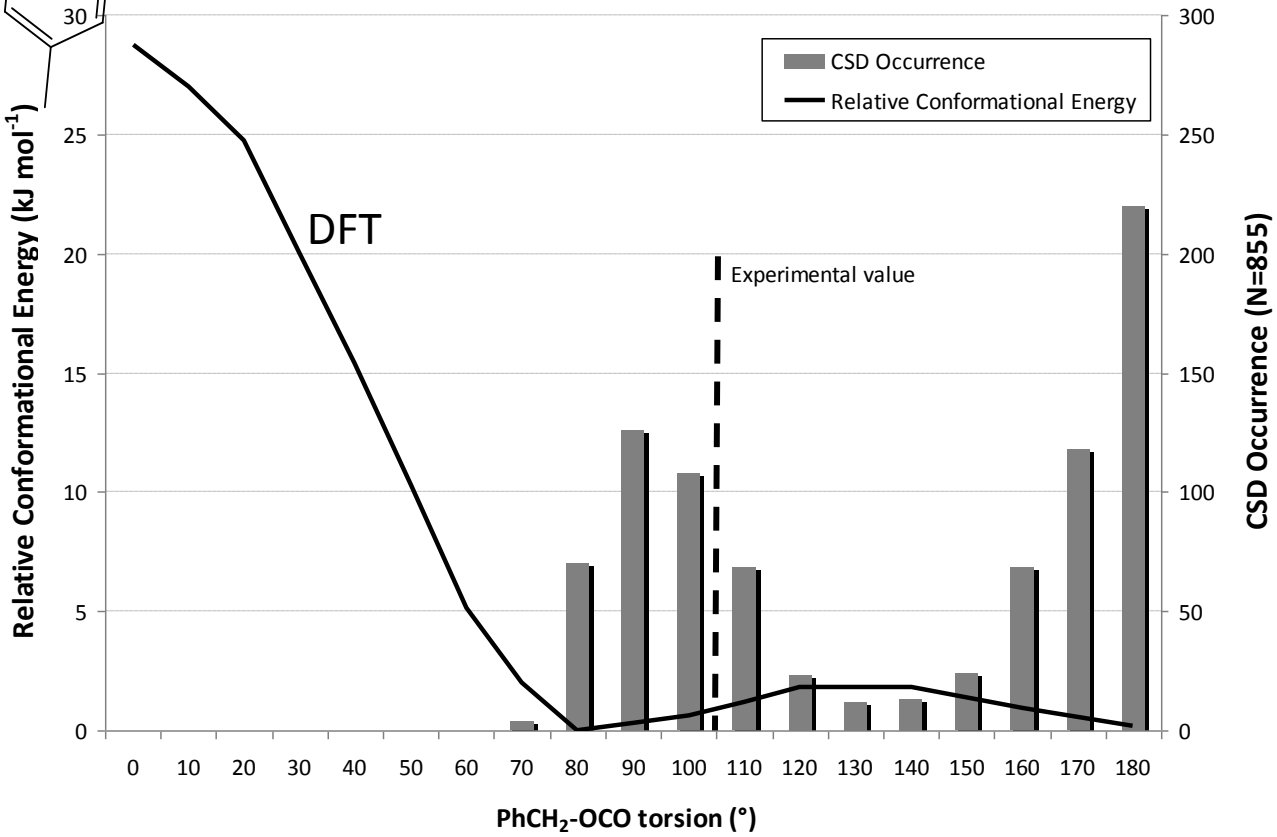


CH2-O torsion



Exploring conformational space:
Data mining (CSD) vs QM conformational energy calculations



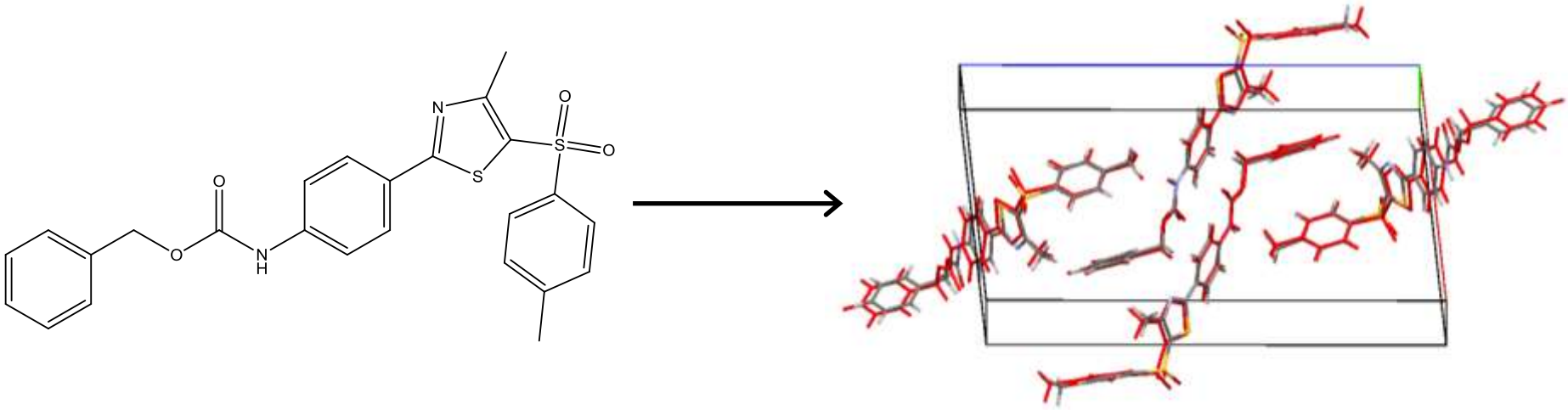


Database information allows a quicker assessment of conformational preferences.
This is energetic information.

Latest 2010 blind test

Similar results to 2007 on small molecules. Publication in preparation.

CSP2010 also included more challenging targets:

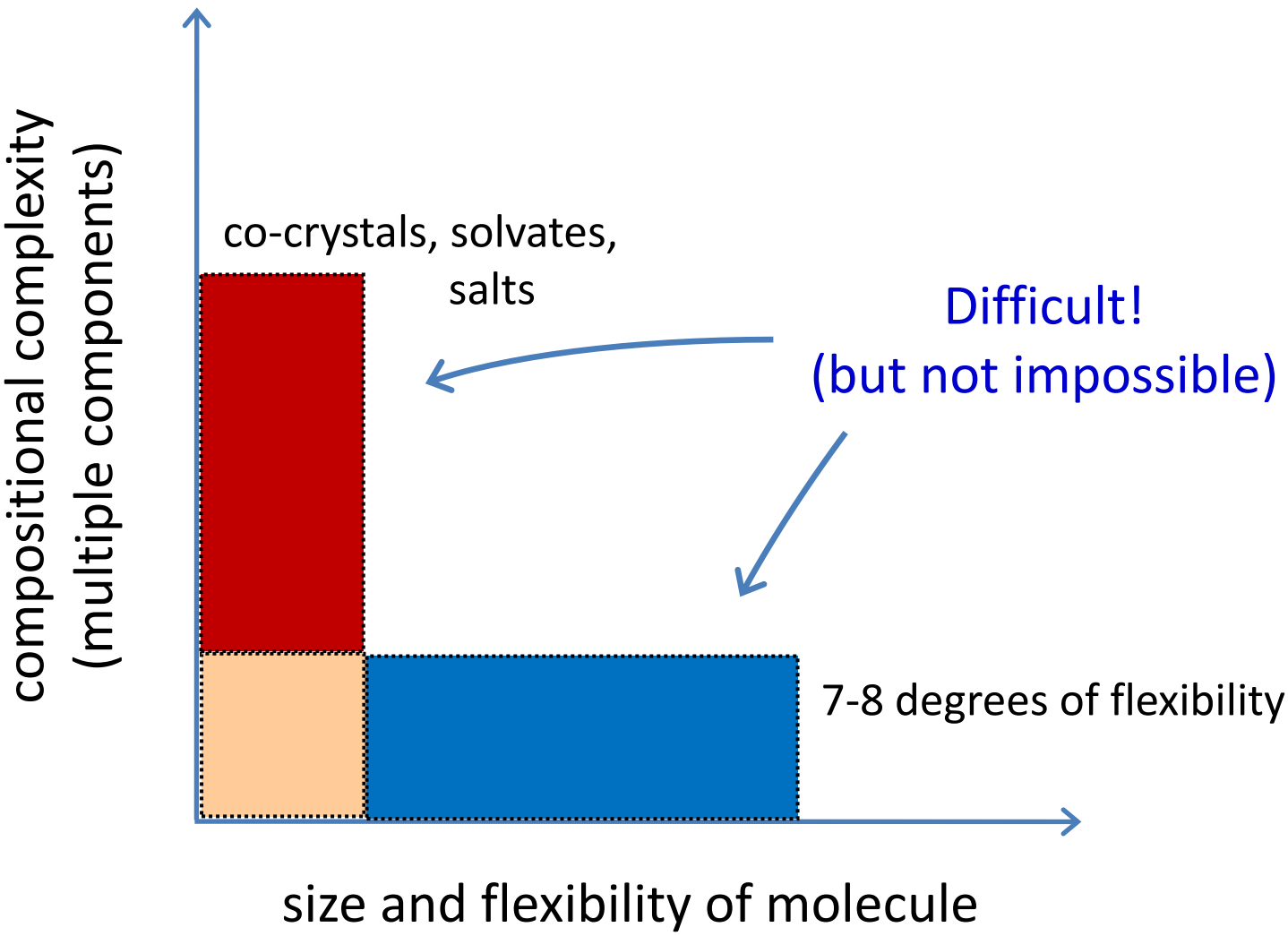


2 groups got this structure correctly, as #1 prediction.

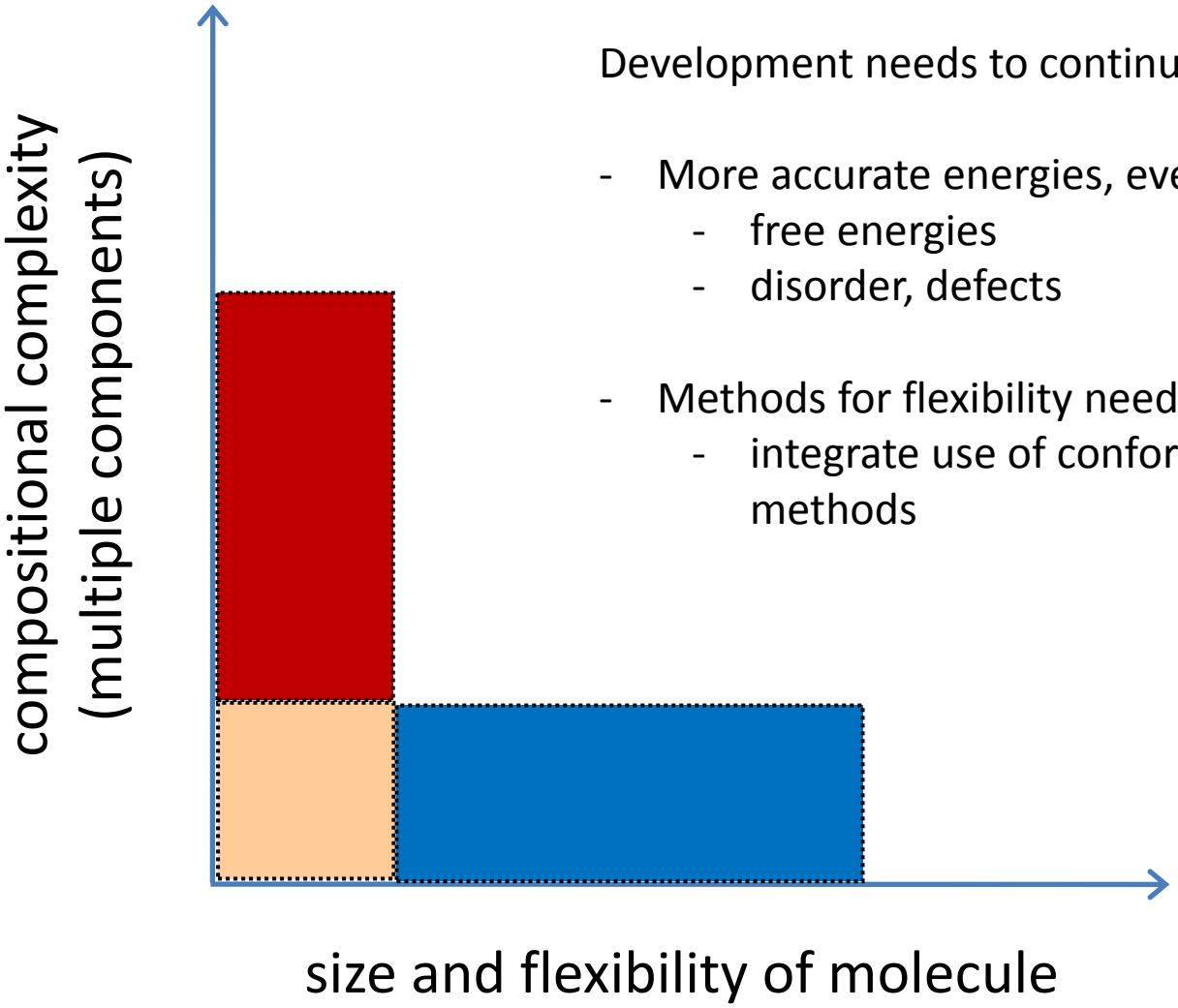
An exciting result: prediction is possible for molecules of this size & flexibility.

However, the low energy structures of these two groups differ significantly, demonstrating remaining uncertainties in the overall energy landscape and possible polymorphism.

Int. J. Pharmaceutics (2011), in press, doi:10.1016/j.ijpharm.2011.03.058.


compositional complexity
(multiple components)
co-crystals, solvates,
salts
Difficult!
(but not impossible)
7-8 degrees of flexibility
size and flexibility of molecule

Looking forward



Development needs to continue in all directions here.

- More accurate energies, even for simplest systems:
  - free energies
  - disorder, defects
- Methods for flexibility need automating
  - integrate use of conformation searching methods

Conclusions

- Crystal structure prediction by lattice energy minimisation has progressed a long way over the past decade.
- These are powerful tools for exploring solid state diversity.
  - guiding discovery of new polymorphs
  - methods can be used to assess possibilities for solvate or cocrystal formation, even where composition is not known
  - developments are also encouraging for flexible molecules
- There is still a lot to do:
  - efficiency of calculations
  - accuracy and reliability of predictions

Acknowledgements

research group:

Dr Tim Cooper
Dr Aurora Cruz Cabeza
Dr Katarzyna Hejczyk
Daniele Tomerini
Hugh Thompson
Edward Pyzer-Knapp
Andreas Stegmüller
Eloisa Angeles Tactay
Jonathan Birtwell
Sebastian Parker

collaborations:

Prof. Bill Jones (Cambridge)
Dr Tomislav Friscic (Cambridge)
Dr Shyam Karki (Cambridge)
Dr Laszlo Fabian (CCDC)
Dr Sam Motherwell (CCDC)
Dr Neil Feeder (Pfizer, Sandwich)
Dr Yuriy Abramov (Pfizer, Groton)
Prof. Sally Price (University College London)
Dr Panos Karamertzanis (Imperial College London)

PPXRD-10 organisers

THE ROYAL SOCIETY

AstraZeneca

The Pfizer Institute for Pharmaceutical Materials Science