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
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.

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

Polymorphism
pharmaceutical implications

Crystal engineering / materials design

Property prediction
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

structural parameters
(molecular orientation, position, lattice parameters)
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 $\sim 10^4$ distinct structures.
An easily predictable landscape:

A typical energy landscape:
This is a big challenge for computational methods:

- $E_{\text{latt}} = -112.91 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -114.72 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -113.58 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -113.37 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -112.69 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -114.89 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -112.59 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -112.23 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -111.63 \text{ kJ mol}^{-1}$
- $E_{\text{latt}} = -111.38 \text{ kJ mol}^{-1}$
approaches to calculating energies:

1) **Atom-atom model potentials**

Typically, an intermolecular model of the form:

\[
U_{ik} = A^{ik} \exp\left(-B^{ik} R_{ik}\right) - C^{ik} R_{ik}^{-6} + U_{\text{electrostatic}}
\]

\(U_{\text{electrostatic}}\) 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?

Presented 5 of the lowest energy calculated structures to ~ 50 crystallographers to visually inspect and choose their “favourite”.

(IUCr, Florence, 2005)
structure A

structure B

structure C

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
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 $\Delta E$ as a measure of success: how far in energy is the “real” structure from the lowest energy predicted structure.

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

$$\Delta E = -4.4 \text{ kJ/mol}$$

$$\Delta E = +1.1 \text{ kJ/mol}$$
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

Results on 50+ small, rigid organic molecules.

Systematic improvements to energies improve success.
- model potentials
  - parameterisation
  - electrostatics
  - polarisation
- QM methods
- free energies: lattice dynamics, disorder

Guiding the experimental discovery of new polymorphs

To the lab... varying crystallisation conditions

Simulated from known DHCBZ structure

DHCBZ grown from ethanol

New peaks

DHCBZ grown from DMSO

DHCBZ ground in DMSO

New form

Simulated XRPD from predicted structures

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

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

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

DHCBZ ground in DMSO

New form
We now want to be extending the range of systems that we can study.
Co-cry stallisation & solvate formation

Introducing a second molecular component can tailor properties

eg. paracetamol (poor compressibility)

Questions that we should ask of computational methods:

• 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:
\[ \Delta H^\circ_{\text{vap}} = 51.6 \pm 1.5 \text{ kJ mol}^{-1} \]
lattice energy (calc) = -58.1 kJ mol\(^{-1}\)

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

More challenging: 2-component structures of unknown composition.
i) Predict all possible crystal structures at a range of stoichiometries
   1:0 (neat crystal); 1:1; 1:2, etc. (M:AcOH)

ratio 1:1

\[
\text{urea:AcOH}
\]

-136.5 kJ mol\(^{-1}\)

ratio 1:2

\[
\text{urea:AcOH:AcOH}
\]

-199.5 kJ mol\(^{-1}\)

ratio 1:3

\[
\text{urea:AcOH:AcOH:AcOH}
\]

-254.3 kJ mol\(^{-1}\)

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 constant composition

no co-crystallisation: \(1 \cdot E_{\text{latt,global min}}^M + 2 \cdot E_{\text{latt}}^{\text{AcOH}}\)

1:1 co-crystallisation: \(1 \cdot E_{\text{latt,global min}}^{M:\text{AcOH}} + 1 \cdot E_{\text{latt}}^{\text{AcOH}}\)

1:2 co-crystallisation: \(1 \cdot E_{\text{latt,global min}}^{M:2\text{AcOH}} + 0 \cdot E_{\text{latt}}^{\text{AcOH}}\)

Acetic acid (AcOH)

Observed as 1:2

Two forms known: 1:1 and 1:2

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

Acetic acid (AcOH)

urea

caffeine

theobromine

Observed as 1:2

Two forms known: 1:1 and 1:2

Predicted to form 1:1

PXRD from material obtained from LAG
simulated from predicted 1:1 global minimum
after Rietveld refinement

1:1 theobromine : acetic acid Rietveld refinement
Red = global minimum predicted
Blue = refinement to PXRD

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!

![Graph showing number of (quasi)-random structures generated versus number of unique crystal structures (after lattice energy minimisation). The graph includes two lines: one for 2 independent molecules, which takes approximately 2 orders of magnitude more CPU time than the line for 1 independent molecule.](image-url)
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!
The diagram illustrates the relationship between size and flexibility of a molecule and its compositional complexity. It shows that flexible molecules are characterized by their size and can form co-crystals, solvates, and salts. Additionally, it highlights that rigid molecules have a lower compositional complexity.
Dealing with molecular flexibility
Now we need to consider all possible conformations.
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

Competing hydrogen bond motifs

Contour lines each represent 4 kJ mol\(^{-1}\)
Hierarchical approach for flexible molecules

1) Crystal structure search
   Rigid conformations
   simple intermolecular potential

2) Intramolecular geometry refinement
   (molecular mechanics)

3) Hybrid model
   atom-atom intermolecular model
   + QM intramolecular energies

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

Polymorphism?
Imperfect energy model?
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


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:

\[ \text{Chemical Structure} \]

Exploring conformational space:
Database guidance
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:

![Chemical structure](image_url)

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.

The diagram illustrates the relationship between the size and flexibility of a molecule and its compositional complexity (multiple components). Co-crystals, solvates, and salts are represented by the red section at the top, indicating a high degree of compositional complexity and a difficult level of handling due to the multiple components involved. The blue section, which represents a lower degree of compositional complexity with 7-8 degrees of flexibility, suggests that while it is more manageable, it is not impossible to deal with. The diagram underscores the challenges associated with molecules that contain multiple components.
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

**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