Applying structural informatics approaches to pharmaceutical supply chain processes

Andrew G.P. Maloney, Mathew J. Bryant and Ian J. Bruno

The Cambridge Crystallographic Data Centre

Thursday 12th April 2018
The ADDoPT Project

• **Advanced Digital Design of Pharmaceutical Therapeutics**
  – Four year collaboration between government, industry and academia
  – Instigated by the Medicines Manufacturing Industry Partnership and part funded under the **Advanced Manufacturing Supply Chain Initiative**
Digital Design

If we designed drugs like we design airplanes...

“Why has pharmaceutical research and development lagged so far behind other industries in the development and application of simulation and modelling for research and development?”

Digital Design: Molecules to medicines

Product and Process Design

Upstream

Product Performance

Downstream

Materials properties
Manufacturing classification
Processing rules
Quality systems
Surface chemistry
Release profiles
Particle attributes
Formulation

Primary Manufacturing - Secondary Manufacturing

Active Ingredient (API) → Crystallisation Filtration Washing Drying → Milling Blending Compaction Coating → Products

Processes

Design and control of optimised development & manufacturing processes through data analysis and first principle models

Products

Performance
The Cambridge Structural Database

All small-molecule organic & metal-organic crystal structures ever published.

944,328

USOPEZ
Natural product intermediate, crystallised as a cyclohexane solvate

The Cambridge Structural Database

All small-molecule organic & metal-organic crystal structures ever published.

R. P. Wilkie et al., Chem. Comun. (2016), 52, 10747-10750
Crystal structure is important...
Drug product design and development

Molecule → Form → Particle
Making a CSD Drug Subset

• Drug definition taken from the approved drug database of Drugbank.ca

• Generated using InChI strings and the CSD Python API

• 8632 crystal structures representing 785 drug molecules

• Searchable and sortable by categories like hydrates, solvates, salts, co-crystals, pure drug (or any combination of these)
Making a CSD Drug Subset

- Pure Compound: 1693
- Hydrates: 327
- Solvates: 351
- Salts: 2126
- Co-Crystals: 1903

Counts at intersections:
- Hydrates & Pure Compound: 46
- Hydrates & Solvates: 15
- Hydrates & Salts: 85
- Pure Compound & Solvates: 16
- Pure Compound & Co-Crystals: 24
- Solvates & Salts: 293
- Solvates & Co-Crystals: 292
- Co-Crystals: 88

Note: The numbers represent the counts of specific compounds or structures within each category, as indicated in the Venn diagram.
Comparison to organic molecules in the CSD
Comparison to organic molecules in the CSD

Molecular weight analysis

- Drug Subset
- All Organics

<table>
<thead>
<tr>
<th>Molecular Weight</th>
<th>% of subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>200</td>
<td>30</td>
</tr>
<tr>
<td>300</td>
<td>15</td>
</tr>
<tr>
<td>400</td>
<td>20</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>600</td>
<td>5</td>
</tr>
<tr>
<td>700</td>
<td>2</td>
</tr>
<tr>
<td>800</td>
<td>1</td>
</tr>
<tr>
<td>900</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td>More</td>
<td>2</td>
</tr>
</tbody>
</table>

Graph showing the distribution of molecular weights for Drug Subset and All Organics.
Solid form and molecular descriptors

Crystal-Type - Molecular Weight

- Pure
- Hydrate
- Solvate
- Co-Crystal

Molecular Weight

- 0-100
- 100-200
- 200-300
- 300-400
- 400-500
- 500-600
- 600-700
- 700-800
- 800-900
- 900-1000
- More
Which is the stable wall?

The database of walls indicates that A is the frequently observed arrangement and therefore the one that achieves stability.

- **A** - Great Pyramid of Giza c.2560 BC
- **B** - Great Wall of China c.1368
- The CCDC c.1992
- My House c.1988
- Hadrian’s Wall c.122
Crystal Structure Characteristics that Influence Stability

Hydrogen Bond Donor/Acceptor Pairing

Molecular Conformation

‘Non-Hydrogen Bond’ Intermolecular Interactions

Hydrogen Bond Geometry, Symmetry and Motif

We mine the CSD to identify these intra and intermolecular geometric preferences.
CSD derived knowledge bases

Mogul

Molecular geometry distributions
- Bond lengths
- Valence angles
- Torsion angles
- Rings

IsoStar

Intermolecular geometry analysis
- Interaction distributions displayed as scatterplots or contour surfaces
- 18,000 pre defined interaction scatter plots
Understanding conformational complexity
Full Interaction Maps (FIMs)

- IsoStar libraries used to map interaction preferences around complete molecules in a crystal structure
- The satisfaction of the Full Interactions Maps by the packing shell of the crystal structure can then be used to assess stability
Using FIMs to assess stability

N. Blagden et al., Int. J. Pharm. (1998), 172, 169–177
Crystal structure directs...

- Morphology/crystal growth
- Surface chemistry
- Mechanical properties
- Solubility
- Stability
- Melting point
Predicting morphologies

- Start from a base morphology prediction
- Assume nucleation onto existing faces to be the rate limiting step in further crystal growth
- Use a forcefield to quantify the most favourable site of interaction
- As growth rate is proportional to the nucleation rate, this allows us to use nucleation kinetics, including a term for supersaturation

Predicting morphologies

Ritonavir Form-II

Supersaturation

Aspirin
Mechanical properties from structure

**Sulfamerazine**

<table>
<thead>
<tr>
<th>Compaction pressure (MPa)</th>
<th>Tablet Tensile Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>20</td>
<td>1.5</td>
</tr>
<tr>
<td>30</td>
<td>2.0</td>
</tr>
<tr>
<td>40</td>
<td>2.5</td>
</tr>
<tr>
<td>50</td>
<td>3.0</td>
</tr>
<tr>
<td>60</td>
<td>3.5</td>
</tr>
<tr>
<td>70</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**6-chloro-2,4-dinitroanilene**

<table>
<thead>
<tr>
<th>Compaction pressure (MPa)</th>
<th>Tablet Tensile Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>20</td>
<td>1.5</td>
</tr>
<tr>
<td>30</td>
<td>2.0</td>
</tr>
<tr>
<td>40</td>
<td>2.5</td>
</tr>
<tr>
<td>50</td>
<td>3.0</td>
</tr>
<tr>
<td>60</td>
<td>3.5</td>
</tr>
<tr>
<td>70</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Predicting slip planes
Predicting slip planes

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Observed</th>
<th>Dreiding</th>
<th>cvff</th>
<th>compass</th>
<th>CCDC Rugosity tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLFNMA01</td>
<td>020</td>
<td>002</td>
<td>002</td>
<td>002</td>
<td>010</td>
</tr>
<tr>
<td>SLFNMA02</td>
<td>020</td>
<td>020</td>
<td>020</td>
<td>020</td>
<td>010</td>
</tr>
<tr>
<td>260457 (UCECAG03)</td>
<td>001</td>
<td>001</td>
<td>001</td>
<td>001</td>
<td>001</td>
</tr>
<tr>
<td>CITRAC10</td>
<td>002</td>
<td>002</td>
<td>002</td>
<td>002</td>
<td>001</td>
</tr>
<tr>
<td>260456 (UCECAG02)</td>
<td>001</td>
<td>001</td>
<td>001</td>
<td>001</td>
<td>001</td>
</tr>
<tr>
<td>PUPBAD01</td>
<td>10-2</td>
<td>10-2</td>
<td>020</td>
<td>011</td>
<td>10-2</td>
</tr>
<tr>
<td>PUPBAD02</td>
<td>101</td>
<td>10-1</td>
<td>011</td>
<td>020</td>
<td>101</td>
</tr>
<tr>
<td>HXACAN</td>
<td>002</td>
<td>002</td>
<td>200</td>
<td>200</td>
<td>001</td>
</tr>
<tr>
<td>HXACAN01</td>
<td>010</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>010</td>
</tr>
<tr>
<td>DIJVOH</td>
<td>002</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>001</td>
</tr>
<tr>
<td>260455 (UCECAG01)</td>
<td>10-1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>10-1</td>
</tr>
<tr>
<td>ethyl paraben (FEGLEI)</td>
<td>101</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>101</td>
</tr>
<tr>
<td>propyl paraben (DUPKAB)</td>
<td>101</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>101</td>
</tr>
</tbody>
</table>

Making the most of every crystal structure ever published

Molecule  Form  Particle
Acknowledgements

- Mat Bryant
- Mat.Sci. Team at CCDC
- Members of the ADDoPT Consortium