Introducing the CSD Solid Form Suite

Knowledge-based development and risk assessment

- How can data mining help understand the solid form behaviour of an active ingredient?
- How many polymorphs are there for a given active ingredient?
- What is the risk that there is a more stable form?
- How do I find a form with improved properties?

CSD Solid Form Suite has been developed to address these and other questions, using a Solid Form Informatics approach.

**Solid Form Informatics**

Solid Form Informatics supports key decisions about development routes of solid formulations of drugs, agrochemicals and molecular materials. It provides knowledge-based assessment of potential polymorphism and associated manufacturing risks, as well as information vital to reformulation and the tuning of property profiles. Solid Form Informatics distils the vast amount of structural data available today into targeted and actionable information. It is poised to become a powerful new asset supporting a Quality by Design approach and its adoption can help significantly reduce the risk of late stage failure that has been so costly to the industry.

**The cost of failure**

If there are gaps in the understanding of polymorphism and the solid form behaviour of a material, the consequences can be severe. Late stage failure is extremely costly and problems encountered after launch can be devastating. In the widely reported case of ritonavir, Abbott experienced manufacturing problems and were no longer able to produce the original form of the drug. The human costs started to mount as the drug was in wide use as part of antiretroviral treatment. The case was finally resolved as a more stable polymorphic form was identified, and new production and distribution methods adopted. The case cost Abbott hundreds of millions to resolve, and an estimated $250m in lost sales in 1998 alone. Another well known case is that of the early-onset Parkinson’s disease drug rotigotine. Rotigotine was filed in 2003 as a substance that doesn’t show polymorphism and was administered successfully as ‘Neupro’ skin patches. However, in 2008 dendritic structures were observed in these patches. A new form had crystallised and this affected the efficacy of the patches. The product had to be withdrawn, affecting the improved quality of life many patients had experienced from the sustained release patches.

**The Crystal Form Consortium**

In 2008, a number of leading pharmaceutical and agrochemical organisations partnered with scientists from the CCDC to address the lack of predictive analytics tools for solid form development. The aim was to develop new tools to inform on polymorphism and aid co-former design, based on the knowledge contained in the millions of intermolecular interactions stored within the Cambridge Structural Database (CSD). Guided by member priorities and CCDC expertise in data mining, the Consortium developed Solid Form Informatics tools, focusing initially on polymorphism.

**Polymorphism assessment**

The result of the Consortium efforts is the CSD Solid Form Suite. In particular, the suite enables knowledge-based polymorphism assessment based on a statistical analysis of hydrogen bonding patterns. The new H-bond Propensity tool predicts hydrogen bonding outcomes, using the knowledge of intermolecular interactions contained within the CSD. By understanding how the environment of a functional group affects its ability to participate in intermolecular interactions, it is possible to predict H-bonding outcomes in target structures. Relevant structures are abstracted from the database and used to build a predictive model. Values for the propensity for a hydrogen bond to form between two functional groups can then be calculated. High propensity values indicate likely intermolecular interactions, whereas low propensity hydrogen bonds can indicate issues with stability.

Combining these propensities for hydrogen bond formation with a statistical model that captures information about how often a functional group participates in an H-bond allows the in-silico generation of chemically reasonable alternative crystal forms. The resultant view of the solid state landscape of an active ingredient addresses questions such as how likely polymorphism is and whether there is the possibility of a more stable form.

Consider temozolomide (Figure 1), used to treat astrocytoma, and marketed under the brand name Temodal by Schering-Plough. The calculated propensity/participation chart (Figure 2) shows that the Form II crystal structure (indicated by a pink dot) is ranked well on both propensity and participation axes.

Figure 1. Temozolomide, marketed under the brand name Temodal by Schering-Plough.

Figure 2. The propensity/participation chart for temozolomide provides an easily interpretable view of the solid form landscape aiding both understanding and communication.

continued overleaf
Deposition of structure factors

Structure factors are now being accepted along with CIFs for deposition to the CSD via our web-based deposition form at: http://www.ccdc.cam.ac.uk/services/structure_deposit

Crystal structure coordinates are only an interpretation of the underlying electron density. The central storage of the associated structure factors alongside the CIF provides a more complete, unchanging record of the experimental data. With the availability of structure factors, the crystallographic community will be able to return to previously refined structures and re-process them using up-to-date refinement techniques. An example is the PLATON/SQUEEZE methodology for modelling disordered solvent or counter-ion moieties. Published crystal structures could be revisited using this methodology to produce alternative interpretations of the structural model.

“The deposition of structure factors to the CSD is extremely welcome and will greatly facilitate the work of reviewers” said Sylvain Bernès, co-editor Acta Crystallographica Section E. “Archived structure factors also promise to be a rich source of data for research related to statistical bias in x-ray structures and in the detection of systematic errors in data collections”. The deposition of structure factors will also help guard against fraud. Recently there have been a number of high-profile retractions of papers from the peer-reviewed literature. These cases have typically involved manual alterations to cell constants and atom types to produce what appeared to be genuine structure determinations of new compounds. Many of these problems could not have been discovered without the availability of structure-factor files and the International Union of Crystallography (IUCr) now strongly encourages the provision of machine-readable CIF and structure-factor files for every submitted structure. Further information on publication standards can be found on the IUCr website: http://www.iucr.org/index.html/leading-article/2011/2011-06-02

Dr Gary Battle, Marketing and Communications Manager

A Free New Tool for Automated Reduced Cell Checking

A common problem during single crystal x-ray diffraction experiments is the unnecessary collection of datasets for crystals that subsequently turn out to be starting materials or a reaction by-product rather than the intended chemical sample. This can mean many wasted hours of diffractometer time collecting data to determine a crystal structure that is already known and published. The problem can be avoided by checking the unit cell of the sample against the CSD prior to the collection of a full dataset. A reduced unit cell search against the CSD will quickly and accurately identify if there are any publically available structures with similar cells.

A new tool called CellCheckCSD has now been developed specifically for the purpose of performing automated unit cell checks against the CSD during data collection. This program comes pre-packaged with the reduced cell information in the CSD as a stand-alone installer. CellCheckCSD runs as an independent command-line tool which means that the full CSD System does not need to be installed on the diffractometer PC and the user does not even need a CSDS licence.

CellCheckCSD was developed as part of a collaboration with Agilent Technologies to provide the search in an automated fashion through the program CrysAlis®. This automated nature means that as soon as a unit cell has been determined (or updated) within CrysAlis®, the software can perform a unit cell check without any user intervention. Any CSD matches found can be easily viewed within CrysAlis® or output to WebCSD or Mercury for further visualisation.

How to get it

CellCheckCSD is a free service that has been developed for automated use. To download an installer for CellCheckCSD free of charge, users simply need to complete a basic registration step on the CCDC website: http://www.ccdc.cam.ac.uk/free_services/cellcheckcsd/

Dr Pete Wood, Research and Applications Scientist

www.ccdc.cam.ac.uk
A major highlight of the CCDC calendar is the triennial IUCr Congress. This is always an excellent event for CCDC to present our ideas and developments, but most importantly, it’s a fantastic opportunity to listen to our data depositors and users of the CSD system and CCDC software.

CCDC were present in Madrid in force, giving our user community the chance to put our developers, database editors, user support scientists, research scientists and management on the spot! For the first time at an IUCr meeting, we held a CSD Open Discussion Forum. For 90 minutes, the 50 attendees raised a wide variety of topics. Strong support was shown for our decision to accept structure factors for archive and our strategy to first encourage their deposition, and then create a position where the deposition of structure factors along with coordinate data becomes the norm. This was reinforced by our user survey: Although only 16% of depositors at the forum had deposited structure factors, 80% would do so in the future.

We learned of the desire for additional information to be captured in the CSD, the need to be clearer about our editing process, how we complicate structure analysis by changing atom labels which can’t then be altered, and received many suggestions for CSD system improvements. We can’t address everything immediately, but are working hard on new internal informatics systems and a new format for the CSD itself. A lively discussion on the importance of supporting Apple Macs, tablet devices and smartphones elicited some diverse views. We think it best to devote as much of our resource as possible towards the science required to improve the CSD system, but will keep supporting as wide a range of platforms as we can. We will also hold a series of webinars, addressing areas where users felt they would benefit from more training.

CCDC also made a significant contribution to the scientific program. Pete Wood co-organised a successful workshop for younger scientists, while Ian Bruno, Gary Battle and Colin Groom co-chaired a microsymposia on Validation, Error Detection and Fraud Prevention, Applications of Crystal Structure Information in Chemical Education, and Crystallography in Disease and Therapy, respectively. CCDC staff also gave six presentations, covering crystal structure prediction (Aurora Cruz-Cabeza), Uses of crystal structure information in chemical education (Gary Battle and Frank Allen), Structure validation (Matthew Lightfoot), Polymorphism of pharmaceutical materials (Pete Wood) and Business models for database organisations (Colin Groom).

A conference highlight was undoubtedly the Celebration of the life of Lodovico Riva di Sanseverino, founder of the Erice Crystallography Schools. The session was organised by Paola Spadon, Frank Allen and Keiichiro Ogawa and attended by Lodovico’s son and daughter in law, Clemente and Alessandra.

The session began with reminiscences of Lodovico’s career, including several periods spent at the embryonic CCDC in the 1960s and 1970s. It was a delight to hear recollections of these periods from Frank Allen, and also to hear how the Erice schools influenced one of our research scientists, Aurora Cruz-Cabeza.

Lodovico always encouraged music and singing ‘mixers’ at Erice, so the session was magnificently rounded off with six favourite songs from the Erice Songbook, amid a few tears, many smiles and a good deal of laughter.

Come and find us on Facebook and Twitter!

Since February, the CCDC has been on Facebook and Twitter, using both social media platforms as a way of keeping our followers up to date with our latest news and developments. We tweet and post status updates about our software releases and updates, conferences that we are attending, interesting research articles and structures, as well as news about the CCDC staff! If you’re not following us already, you can find us on Facebook at www.facebook.com/ccdc.cambridge and on Twitter at www.twitter.com/ccdc_cambridge.

When we’ve reached 500 followers on Facebook we’ll be giving away a 16GB Apple iPad to one lucky follower, so what are you waiting for? Come and get involved in our online community!

Lauren Thomas, Account and Marketing Manager

www.ccdc.cam.ac.uk
**Spotlight on Relibase 3.1**

Remind me again what Relibase+ is used for? Relibase+ allows you to perform powerful searches and analyses of protein-ligand interaction space. This is particularly useful during project initiation when you quickly need to understand the intricacies of a new protein target. Many people also use it to derive general knowledge about protein-ligand interactions.

I use the PDB to look at protein structures. What advantages does Relibase+ offer? One of the most powerful aspects of Relibase+ is its ability to search for similar cavities in a sequence independent manner. This can be used to generate chemical starting points from parts of ligands bound to sub-pockets similar to the sub-pocket of interest. It can also be used to get a handle on target cross-activity.

That sounds extremely useful. Does Relibase+ offer anything else? Relibase+ offers lots of things. The ability to organise your searches into hit-lists and to combine these using boolean logic. There is also an easy to use sketcher interface for searching for protein-ligand interactions. Both 2D and 3D constraints can be added to the sketcher searches as well as protein secondary structure constraints. The list goes on. However, one of the most important features of Relibase+ is that it allows you to create your own databases to archive your in-house structures. These can be searched either independently of or in conjunction with the data derived from the PDB.

I remember struggling to create in-house databases when using Relibase+ in the past... You and many others... This is why, in version 3.1, we have been working on improving the user experience of creating in-house databases. Both the web-based graphical user interface as well as the command line interface has been overhauled. The graphical user interface makes it trivial to create small databases of handfuls of structures, or large databases if updated continually as new structures are determined. The command line interface is more powerful and makes it possible to port large backlogs of in-house protein structural data, minimising the amount of manual editing required.

That’s great! I have been looking for an easy way of storing and organising my in-house structures. When will version 3.1 be available? Version 3.1 of Relibase+ was released in June so you can access it now. Please get in contact with admin@ccdc.cam.ac.uk.

Wait a minute, I remember Relibase+ being really difficult to install... Not to worry. We have also overhauled the installer in version 3.1. It is now trivial to install Relibase+.

Dr Tjelvar Olsson, Research and Applications Scientist

---

**Events**

Over the coming months you can come and meet us at various meetings and events around the world.

<table>
<thead>
<tr>
<th>Date</th>
<th>Conference, meeting or event</th>
<th>Venue</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-9 February 2012</td>
<td>Lorne Conference on Protein Structure</td>
<td>Lorne, Australia</td>
<td>Attendance</td>
</tr>
<tr>
<td>25-29 March 2012</td>
<td>Spring ACS Meeting</td>
<td>San Diego, CA, USA</td>
<td>Talk, Exhibition</td>
</tr>
<tr>
<td>16-19 April 2012</td>
<td>Spring BCA Meeting</td>
<td>Warwick, UK</td>
<td>Exhibition</td>
</tr>
<tr>
<td>10-15 June 2012</td>
<td>Gordon Research Conference on Crystal Engineering</td>
<td>Watervale, VH, USA</td>
<td>Attendance</td>
</tr>
<tr>
<td>17-20 June 2012</td>
<td>Fourth European Conference on Crystal Growth</td>
<td>Glasgow, UK</td>
<td>Attendance</td>
</tr>
<tr>
<td>15-18 July 2012</td>
<td>SLA Annual Conference &amp; INFO-EXPO</td>
<td>Chicago, IL, USA</td>
<td>Attendance</td>
</tr>
</tbody>
</table>

---

**CCDC Publications**

The CCDC team frequently publish results of their research, which is often the work of collaboration with industrial or academic scientists. You can find the full list of our publications at [www.ccdc.cam.ac.uk/publications](http://www.ccdc.cam.ac.uk/publications).

Here are our most recent titles, published since May 2011:

- The extent of enthalpy-entropy compensation in protein-ligand interactions (2011)
- Designing a new Diels-Alderase: A combinatorial semi-rational approach including dynamic optimization (2011)
- Structure prediction, disorder and dynamics in a DMSO solvate of carbamazine (2011)
- Docking performance of fragments and drug-like compounds (2011)
- Deducing chemical structure from crystallographically determined atomic coordinates (2011)
- Teaching 3D structural chemistry using crystal structure databases 3: The Cambridge Structural Database System – database content and access software in educational applications (2011)
- Teaching 3D structural chemistry using crystal structure databases 4: Advanced examples of discovery-based learning (2011)
- Accelerating Molecular Docking Calculations Using Graphics Processing Units (2011)
- The Effect of Pressure on the Crystal Structure of Bianthron (2011)
- The Effect of Pressure on the Crystal Structure of Bichromate (2011)

Using the outer coordination sphere to tune the strength of metal extractants (2011)

---

**Over the coming months you can come and meet us at various meetings and events around the world.**