50 Years of Sharing Structures

Opening lecture by Olga Kennard, Founder of the CCDC

- Ian Bruno (CCDC)–"Sharing research data and knowledge –a fifty year perspective"
- Helen Berman (PDB)–"The evolution of the Protein Data Bank"
- Suzanna Ward (CCDC)–"A journey through the Cambridge Structural Database

Report by Hannah Bruce Macdonald, University of Oxford

Every seat was filled for the opening talk of the 50th anniversary of the Cambridge Structural Database (CSD): **Olga Kennard**, the first Director and one of the founders of the CCDC in 1965. She began by talking of JD Bernal and his belief that huge scientific gain could be made by bringing together the results of many individual experiments, and how this idea lay at the foundation of the CSD. Looking back at how one post-doc would process 100 structures onto punch cards using knitting needles gave context as to how far the information had come. The Royal Society saw the early work and Olga was invited to present their work in Washington, which led to the funding that allowed for one post doc and the beginning of the CSD. She accounted the success to the talented and energetic group who started the daunting task of work with the database, to whom users of the CSD owe so much. Olga talked of being offered an existing database, which hadn't been checked but this was turned down, as she wanted to build a reliable database, which people could use with confidence. She closed by telling the room how 'the basic ideas still hold good' and how much of the dream had been realised. Her talk of the history of the Cambridge Crystallographic Data Centre (CCDC) illustrated the highlights of the 50 years she has seen of the company, setting the scene for the talks of the scientific successes and working to come, which have all been made possible by the database.

Ian Bruno, a senior manager at the CCDC followed, giving a 50-year perspective of the CCDC. He talked of the flood of information the CSD faced, and the methods used to handle it. He discussed the data formats used to help that – the introduction of the CIF files, and the positives and the negatives of human-readable file formats. The Public Library Of Science now expects all data to be submitted with a journal submission, and Ian described how this is already compatible with the CSD system. He debated the addition of raw data to the database, and how this could be beneficial with improvements to analysis techniques allowing information to be revisited, but came at a storage cost. He closed, talking of the licences available for the database, and how no institution would be denied access to the service based on funding.

Helen Berman, former Director of the Protein Data Bank talked about the PDB and how the CSD was instrumental to its development. She attended the 1971 Cold Spring Harbor Symposium where the PDB was born through a petition. When Walter (Hamilton) established the archive, the first step he took was to fly to England and

discuss the CSD with Olga. The PDB was made possible by recognising the importance of the CSD in its field. The PDB now receives 210 new entries a week, from a broad community. She echoed Ian's discussion of raw data, stating how despite the benefits the funding and resources are limited. The parallels between the two databases are now reflected nicely as they now both have offices in the same building in New Jersey.

Suzanna Ward, Cambridge Structural Database Manager at the CCDC, closed the first session of the Symposium, discussing the numbers of the CCDC, 784,428 was the current number of structures in the database, and how the R-factor had decreased over time, and looked at how publications were now split between many more journals. There are now many more authors per crystal structure, and the average cell volume has now doubled in size. The increase in size and complexity in the structures reflects the improvements in techniques, and there are very few elements left, that are not present in a CSD structure. She discussed the alphabetical bias by the users of the database, with a 7% increase in looking at early alphabetical structures. She finished by discussing some of her favourite structure refcodes, including CARPET, BIKINI and BADBOY.

50 Years of Science & Software

- Robin Taylor (CCDC) – "CSD research at the CCDC: a voyage through the years"
- Angelo Gavezzotti (University of Milan) – "Twenty-five years of Cambridge Structural Database mining: Chemical bonds and chemical bonding"
- Jason Cole (CCDC) – "The development of the CSD System: The challenges faced and the milestones achieved"

Report by Andy Maloney, University of Edinburgh & the CCDC

Having woken up at the crack of dawn to endure the hustle and bustle of an early morning flight to get from Edinburgh back to my CCDC base of operations (and being only slightly delayed), I was extremely glad to arrive at the tranquillity and grandeur of Downing College. A truly fantastic setting for what promised to be a truly excellent conference.

After a nice spot of lunch in the sun spent catching up with some old friends and some fascinating talks discussing the journey of the CSD so far, of particular note the opening lecture from Olga Kennard, it was time to settle down to the late afternoon session – "50 Years of Science and Software". Our session chair, Ian Bruno, took to the stage to introduce the first speaker, the CCDC's own **Robin Taylor,** and his talk "CSD research at the CCDC: A voyage through the Years". During his introduction, Ian was quick to point out the caveat in Robin's abstract: "Only a fool would attempt to summarise these fifty years of CCDC research in just thirty minutes." How would our speaker fare, having made such a rod for his own back?

Robin began by reminding us why research is done at the CCDC – to develop our software and understand our users, to maintain our contacts and our own high profile and, of course, because we are scientists and simply enjoy doing it. He noted that the research at the CCDC is becoming more and more diverse with each passing year before taking us on a timeline of a set of papers that illustrated this point magnificently, and highlighted how important the CSD and the research the CCDC provides are to the scientific community. All the famous papers from through the years were there, from the evidence of C-H···O hydrogen bonds to the tables of derived bond lengths (cited over 12,000 times!) and the introduction of the Crystallographic Information File, and many more besides. More recently, last year's paper "Knowledge-based approaches to co-crystal design" showed how far the CSD and the CCDC have come in terms of harnessing the huge amounts of data at our disposal to tackle complex problems, highlighting the importance of several high-profile collaborations along the way. Clocking in at just under half an hour, Robin had summed it all up perfectly, foolishly proving himself wrong in the process.

Our next speaker was **Angelo Gavezzotti** of the University of Milan, with his talk titled "Twenty-five years of Cambridge Structural Database mining: Chemical bonds and chemical bonding". After a historical preamble, he challenged the audience with two questions. The first of these was to ask if we know what a chemical bond is.

Fortunately, courtesy of the CSD, we do (to some extent anyway). The second question was a bit trickier. “What is **not** a chemical bond?” Angelo went on to stress that, while the distribution of intramolecular bond lengths across the CSD is quite narrow, intermolecular bond lengths for hydrogen bonds and “the sons of a lesser god” (other short contacts, to you and me) have considerably wider distributions. A stark warning, perhaps, that it is very important to investigate energies as well as geometries before making any assertions about bonding.

This session was rounded off by **Jason Cole**, a member of the CCDC’s staff ever since completing his PhD, with his talk, “The development of the CSD System: The challenges faced and the milestones achieved.” It must be said, Jason was ideally suited to give such a presentation, having contributed to the majority of the software involved. I have to admit, although I’ve used the CSD system almost every day for the last five years, I hadn’t ever really thought about the journey it had taken. From the hefty tomes of the early days which had to be pored through manually, to the sleek searches that can be performed in the blink of an eye today, the CSD System has come a long way. Jason spoke with great insight into how the changing scientific world has led to numerous data explosions over the years that the CSD has had to cope with.

And cope it has. The CSD, through some pieces of remarkably clever software, has always managed to stay ahead of the curve. Perfectly summed up by Professor Gavezzotti, “the CSD is to the structural chemist what lavender is to the bumblebee.”

Molecular Recognition

- Martin Stahl (Roche) – "Mining the treasure trove: Interaction and conformation searching in structural databases"
- Chris Hunter (University of Cambridge) – "Quantification of non-covalent interactions"
- Gerhard Klebe (University of Marburg) – "From structure correlation in the CSD to the prediction of molecular recognition in protein-ligand complexes"

Report by Christin Schärfer, CCDC

The first speaker of this session was **Martin Stahl** who just recently joined the board of the CCDC's trustees. He started his talk "Mining the treasure trove: Interaction and conformation searching in structural databases" by explaining how he became interested in chemical structures. As a child he collected stamps, some of them with pictures of molecules. He really enjoyed it but later realised that it is much more fun to look at structures than at stamps and he started wondering what people know about conformations and how we can share knowledge. Martin suggested that in order to expand our knowledge about conformations we should use analogies between structural motifs and we should think in series of structures rather than individual cases. In an application example he explained how they successfully used this methodology at Roche to look at conformations of Suvorexant to find out what the overall shape looks like. CCDC tools like Conquest were really helpful during this process. In the second half of his talk Martin described and showed applications of a new pharmacophore query tool that has been developed in collaboration with the CCDC. Results are provided almost in an instant by the new tool, allowing iterative searches and making them highly interactive. Martin finished his talk by showing more examples that illustrate how crystal structure data can be used to gather knowledge that helps analysing structures.

The next talk "Quantification of non-covalent interactions" was given by **Chris Hunter**. He introduced himself by saying that although he shares a birthday with the CSD he is more a solution guy. To start his talk Chris showed a slide with his first contribution to the CSD which is probably also the last molecule he ever made with his own hands. Chris's group perform quantitative measurements of the thermodynamic properties of aromatic stacking, hydrogen bonding and halogen bonding interactions in the liquid phase. They use IsoStar and quantum chemical calculations in the gas phase to corroborate their findings. Chris established a system to use interaction potentials for functional groups derived from the liquid phase experiment to screen for compounds that form co-crystals with a particular drug molecule. This co-crystal prediction project is done in collaboration with Neil Feeder at the CCDC. This approach simplifies complex systems by approximating the overall energy of association as a sum of individual interaction energies, which are rigorously benchmarked against experimental data. Klaus Müller asked whether the model can explicitly model the cooperative effects of weak interactions. Chris answered that it cannot, but that it is rather accurate nonetheless.

The session was finished by **Gerhard Klebe** and his talk “From structure correlation in the CSD to the prediction of molecular recognition in protein-ligand complexes”. Gerhard is a former CCDC trustee and started his talk by telling us that his first encounter with the CSD was during his PhD in 1979 and a book written by Jack Dunitz. He then talked about his time at BASF where he got acquainted with computational methods for generating conformations. Looking at these conformations they realised that the conformations often didn’t correspond with crystal structures and so they came up with the idea to solve this problem by using torsion angle distributions from the CSD. In his next slides Gerhard showed how they predicted interaction sites in protein pockets by mapping crystal field environments in the CSD which later resulted in the development of IsoStar. Next he talked about a problem that occurred while looking at preferred atom-atom distances in protein ligand complexes for scoring functions. At that time, there were not enough entries in the PDB and the resolution of the entries was not very good. To make the best possible use of all the information present, Gerhard and members of his group including Manfred Hendlich developed Relibase and Relibase+ in collaboration with the CCDC and others. One of Relibase’s very well received features is the ability to store positions of conserved water molecules. Gerhard showed an impressive example of how the quality of the water network in a binding site affects the potency of a drug. He finished his talk by saying that the CSD is a great tool which everyone in Marburg is very thankful for and that their research really depends on the CSD.

Molecular Design

- Klaus Müller (Roche) – "The CSD and Roche's early entry into structure-based drug discovery"
- Terry Stouch (Science for Solutions) – "The CSD: A fundamental resource for molecular modeling"
- Alberto Gobbi (Genentech) – "We need Champagne, other drinks are not enough!"

Report by Florian Roessler, University of Cambridge

The session was opened by the chair, Beth Thomas, who introduced **Klaus Müller** (Roche) as the first speaker of the session. In his talk Dr. Müller told the compelling story of how structure-based drug design at Roche was influenced from the beginning by very fruitful collaborations with the CCDC and the PDB. This is exemplified by the shared 50th birthday of the CSD and Roche's small-molecule X-Ray structure analysis efforts. He continued describing his early career at Roche, where he made access to structural databases a condition of his involvement. While in the early years Roche was using its own relational database version of the CSD (ROCSD), they abandoned this project in the 1990s in favour of the powerful CSD software suite. The early years in the area were made challenging by the lack of sufficient crystal structures and it took a significant amount of time and effort to develop the necessary molecular models needed in their ongoing projects. He illustrated this by drawing the listeners' attention to their efforts in producing sufficiently hinged small-molecule structures in order to target the E. coli DHFR-MTX complex. Their model existed as early as 1982 but it took until 1986 until the first compound (CSD REFCODE: DUZHEL) was produced and showed sub nano-molar activity. Previously found compounds had shown no activity at all. He continued by stating that underrepresentation of conformational polymorphs due to crystal packing effects still exists despite today's large number of structures in the CSD. He summarised his talk by saying that because of the growth of the data in the CSD and the software that is built around it, previous challenging questions can nowadays be answered more easily and elegantly than before.

The second speaker of the session was **Terry Stouch** who works as a consultant with Science for Solutions. Dr. Stouch shared his great insight into the early days of force-field development and his involvement therein. He described the early advances in the field and the close interactions with researchers working on the CSD. As his talk continued he highlighted milestones like the acquisition of energy parameters from crystal data in 1979 and the work by Donald E. Williams on deriving non-bonded potential parameters from crystals. The lack of crystal data representing substantial volumes of chemical space along with the rise in availability of computational resources then led to the emergence of force-fields that were parameterised using quantum-mechanical approaches. While over the past decade these types of force-fields have shaped development in the area he stressed that it has come to a point where crystal data has again become more relevant. He attributed this to the significant improvement in availability and quality of crystal data. As an example he

discussed the dihedral angle of ligands in a structure-based drug optimisation context. In this example, the difference in steepness of the energy profile between CSD and QM dihedrals nowadays can provide valuable information and influence the outcome significantly. The talk was concluded by his emphasis on the significance of the CSD as an educational tool and the need to promote interaction between Modellers and Crystallographers as exemplified in a successful RCSB workshop in 2009 at Rutgers University.

Alberto Gobbi was introduced as the last speaker of the session. Along with the previous speakers, Dr. Gobbi presented a convincing story of the importance of the use of crystal data in a drug development context both in the past and for the future. His talk focused around the importance of the correct assessment of strain energy between bound and unbound ligands and its relevance in determining the strength of ligand binding. He provided insights into cases where the calculation of strain energy using modern force-fields still performs below expectations. In particular considering the time and resources that are involved in providing accurate dihedral parameters, classical force-field approaches are seen as a bottleneck in this aspect of Structure-Based development. He presented examples of the trade-offs that current approaches encompass. While quantum-mechanical approaches at high level of theory provide good insights into the accurate potential energy surface of ligands, the time involved in running these calculations (from 30 min up to 12 hours per dihedral) render them unusable in a high throughput context. On the other hand, force-field approaches while providing a much quicker result, still struggle with a lack of accuracy due to the generality of their parameters. Besides this, Dr. Gobbi also highlighted how force-field based approaches can massively reduce the time spent on drug optimisation problems. In addition to this he stated that only by comparing the relevant calculations to data from small molecules of the PDB (bound) and CSD (unbound) can we ensure that our models provide the best results possible. All this contributes to better tools that with their interactive capabilities significantly improve the understanding of the role of strain energy in protein ligand binding. He concluded his talk and the session by emphasising that champagne and much praise reflect only part of the appropriate way to celebrate the 50 year anniversary of the CSD and its contribution to the scientific community.

Solid Form Informatics

- Susan Reutzel-Edens (Eli Lilly and Company)–"Lessons learned in structure-based solid form design"
- Joel Bernstein (Ben-Gurion University of the Negev & NYU Abu Dhabi)–"The CCDC and me"
- Aurora Cruz-Cabeza (Roche)–"From desmotropy to conformational polymorphs"

Report by Luca Iuzzolino, University College London

This session of talks during the '50 years of the Cambridge Structural Database' event was focused on the role the CSD solid form informatics tools have had in aiding scientific research and the understanding of organic-solid state behaviour both in industry and academia. This series of talks was characterised by a combination of personal experiences, anecdotes and scientific information that made it extremely interesting for the audience.

Susan Reutzel-Edens who works as a Senior Research Advisor for the pharmaceutical company Eli Lilly and Company, gave the first talk. It was focused on how the CSD helped to solve some important drug-development problems during her career at Eli Lilly. A very interesting example she gave regarded the synthesis of pruvanserin, a drug used to treat insomnia: during the development it was realised that its crystals lost weight with increasing temperature, which suggested the presence of water within the crystal structure. Although this intuition was backed by NMR studies, she struggled to make her management believe this theory since there was no way to see the presence of water in the crystal structure. But the CSD informatics tools solved the problem by allowing a change in perspective that made it possible to see the presence of water in the voids between molecules, which would have not been possible without that presence. She also reminded the audience of how the Blind Tests of organic crystal structure prediction organised by the CCDC have increased the credibility of computational methods in industry leading to her collaboration with Sally Price's group at UCL.

The second talk was given by **Joel Bernstein**, currently serving as a Professor at New York University in Abu Dhabi and at the Ben-Gurion University of the Negev. After an interesting anecdotal introduction about how a piece of homework in his youth made him understand the importance of finding the right information when needed, he went on to talk about some major developments in our scientific understanding of solid state organic chemistry that would not have been possible without the presence of the CSD. In particular the fundamental role of the CSD in making the scientific community accept the physical existence of the weak C-H···O bond despite the famous Jerry Donohue's "it isn't" was outlined. Professor Bernstein also expressed his gratitude to the CSD for having made the hydrogen bonding graph sets become lingua franca of chemistry through their implementation into the CSD informatics tools. This was very important to him: the idea of graph-sets had been developed by his colleague Margaret C. Etter before her death in 1992, and he was

grateful to the CSD for having allowed her intuition to be spread out in the scientific community.

The final talk of the session was given by **Aurora Cruz-Cabeza**, who is currently working at Roche in Basel. It was focused on how the CSD has helped her throughout her career as a researcher to investigate some solid-state phenomena. In particular she gave a very interesting exposition of how the use of the CSD allowed her study of how to control tautomerism via supramolecular selectivity. The main section of the talk was focused on her recent study conducted together with Joel Bernstein on polymorphism. In particular mining the CSD with the aid of its informatics tools has allowed them to collect a large enough set of data to be able to demonstrate how certain common beliefs about polymorphism do not have any statistical base: the occurrence of polymorphism appears to be totally independent of molecular flexibility, molecular size and hydrogen bonding. The final part of the talk was focused on how the CSD was vital in categorising and studying conformational polymorphism, which occurs when a molecule crystallises in two different conformers separated by an energy barrier. Data-mining with CSD informatics tools has also made it possible to produce some simple cut-offs to recognise conformational polymorphs.

Overall it was a very interesting session, which gave every person in the audience a very good idea of the importance of the informatics tools developed by the CSD in scientific and industrial research.

Structural Chemistry

- Paul Raithby (University of Bath) – "The use of the CSD in understanding and designing solid-state organometallic reactions"Greg Ferrence (Illinois State University) – "Permeating the Cambridge Structural Database into chemical education"
- Nick Funnell (University of Oxford) – "Disorder and dimensionality"
- Zéphirin Yav (University of Kinshasa) – "CSD use at the University of Kinshasa in D. R. Congo"

Report by Rachael Skyner, University of St Andrews

The structural chemistry session, chaired by the CCDC's own Pete Wood, was kicked off by **Paul Raithby**, Professor of Inorganic Chemistry at the University of Bath. Paul was introduced by Pete as a 'Giant of British Crystallography", and his talk certainly emulated this introduction. Paul's discussion had a focus familiar with the rest of the conference; where have we come in the last 50 years - specifically in structural chemistry? Quoting the experienced words of Jack Dunitz, Paul reminded us all that "Crystals do not contain an array of rigid molecules"; and it is the movement of molecules in crystals that Paul believes to be a future focus of structural chemistry. The movement of molecules in the solid state is key to Paul's research, which focuses on reactions in the solid state.

Paul discussed his lab's tried and tested method of using Christmas tree lights (!) to induce photochemical reactions. This sort of reaction has been known since the late 19th century, with the first example in the solid state being Cohen and Schmidt's light-induced 2 + 2 cycloaddition reactions (1964). Paul has used the topochemical postulate that the reaction process follows the minimum energy pathway, meaning the least atomic movement is the most favourable pathway for preservation of the crystal in a solid-state reaction, to investigate further. Assuming the structure of the product is related to the orientation of the reactant monomer, Paul searched the CSD, finding 67938 hits corresponding to parameters for 2 + 2 cycloaddition, with around 2600 of these containing the necessary parallel double bond. Of these structures, 4 structures were found which hadn't previously been investigated for the photochemical 2 + 2 cycloaddition reaction, which Paul's group went on to investigate. This sort of example reminds us all of just how far and wide the use of the CSD stretches, and of how the potential applications of the vast amounts of data we have at our fingertips are far beyond what many of us would dream of!

The second talk of the session was given by **Greg Ferrence**, Professor of Chemistry at Illinois State University, who has been collaborating with CCDC since 2006, focusing on how to use the CSD in education. In 2004, Greg noticed that the CSD was absent from teaching the principles of chemistry at the undergraduate level, which he thought was particularly absurd. In 2006 Greg surveyed the literature for examples of the use of the CSD in chemistry education, and found 15 mentions of the CSD, of which only 3 were related to the use of it. Greg set to work in collaboration with the

CCDC to select a subset of structures representative of the topics covered in undergraduate chemistry. This set is now openly and freely accessible via the CCDC.

Greg also showed us some examples of the specific modules designed by himself and the CCDC to aid teaching. One particular example that sticks in my mind is using the CSD to search for the existence of the bromonium ion, in order to help students understand the mechanism of Br_2 addition to alkenes. If I had been taught about this mechanism with the aid of the CSD for visualisation, maybe I would have really understood it, instead of sitting at the back of the lecture theatre giggling at "backside attack" - but maybe that's more a reflection of my attitude than my lecturer's teaching style!

The penultimate speaker of the session of the session was **Nick Funnell**, who works as a Post Doctoral Research Associate in Prof. Andrew Goodwin's group at Oxford, and the winner of this year's CCDC Chemical Crystallography Prize for Young Scientists. Nick discussed some of his fantastic work focusing on disorder and dimensionality in three separate systems – an organic hydrate, a framework material and an inorganic nanosheet - all very different materials, yet all very interesting. Nick promised us that he really does use the CSD a lot, even though his presentation focused on the nitty gritty details of how he went about solving the disorder in the materials he discussed. Disorder is something that most crystallographers have to deal with at some point, and the methods that Nick discussed certainly had our minds working overtime on how we could improve our own structure solutions!

The final speaker of the session was **Zéphirin Yav** from the University of Kinshasa, Democratic Republic of Congo. The University of Kinshasa first started its relationship with the CCDC in 2007, when the CCDC granted a CSD license to the Sciences Faculty as part of a collaboration whereby over a 24 month period, academic staff and students were introduced to and trained to use the CSD for both training and learning. In 2013, a research collaboration between CCDC and Kinshasa was established, allowing the sponsorship of a number of students in the area of structural chemistry. The CSD plays an important role in supporting the QM calculations conducted by researchers in Kinshasa.

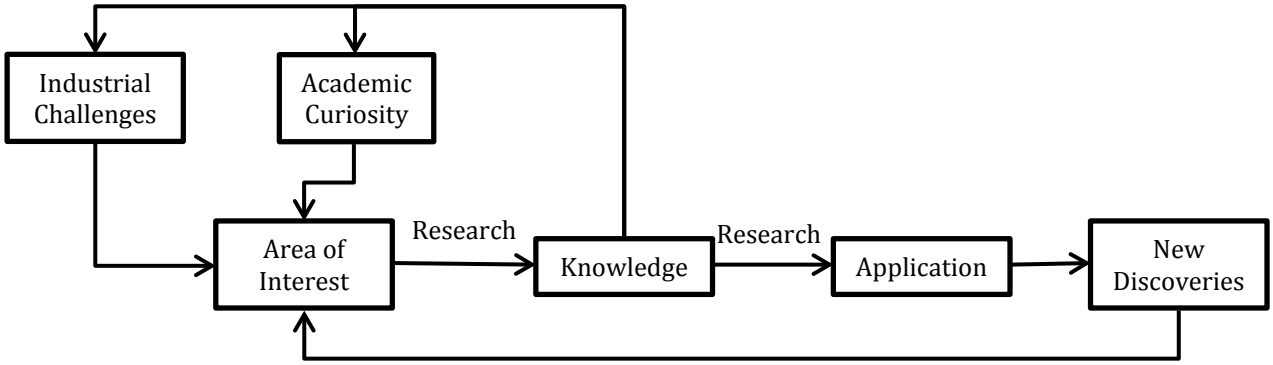
Yav discussed some of the difficulties encountered by his university, and exemplified the responsibility of researchers from top-class institutions to aid the development of research programmes in the rest of the world. When asked "What can we do to help?" Yav simply responded, "Collaborate with us and help us where you can". Certainly food for thought, and a perfect close to the session. We started the session thinking about where we have come, and ended the session wondering where we would go in the future.

Structural knowledge in a changing world

- Chick Wilson (University of Bath) – "From structure to crystallisation and manufacturing: a journey from fundamentals to flow"
- Bob Docherty (Pfizer) – "Towards computational product and process design"

Report by Elena Kabova, University of Reading

It is often stated that academia and industry are very different - and indeed they are. Academia is (historically) largely curiosity driven, whereas industry is driven by economic imperatives. However, there are clear underlying similarities, as demonstrated by the talks of **Prof. Chick Wilson** and **Prof. Bob Docherty**, which focussed on the area of solid-state science in (mainly) the area of pharmaceuticals. Both speakers described their personal journey (very X Factor!), their key drivers and the underlying principles of their work, and these were remarkably similar throughout. My interpretation of these various elements is outlined below:



In Chick's case, the cycle began with sheer curiosity: to discover the fundamentals of the hydrogen bond from a structural viewpoint. This knowledge is then applied to design and create new materials with improved properties. By way of example, his group discovered (with a little bit of serendipity) a crystallisation route for paracetamol form II (Thomas et al., 2011), a form that had previously proved quite elusive. One of his main focus areas is now the continuous flow manufacturing of crystalline forms, which is of course of considerable industrial interest; efficiency, sustainability, reduced production times, decreased costs and quality control. With low-solubility pharmaceuticals increasingly being formulated as co-crystals, salts and solvates, the aspect of applying continuous flow crystallisation to multi-component systems is also under investigation.

Bob then proceeded to show how industry successfully exploits the accumulated knowledge and proactively utilises it in current developments. The importance of the aforementioned hydrogen bond and the CSD, together with its tools, was demonstrated by his referencing of a number of articles - he must have highlighted around 20 papers in his talk. In an early paper of Margaret Etter for example (Etter, 1990), the predictability of hydrogen bonds is discussed, and empirical hydrogen bonds rules are established based on the CSD derived information of intermolecular contacts. This structural knowledge was shown to play an integral role in all stages of

the “molecule to crystal to particle to drug” journey, and helps tackle specific problems, an example of which is polymorph stability (Feeder et al., 2015).

Unsurprisingly, both speakers strongly linked their journeys to developments in the CSD, which in just 50 years has advanced from few hundred simple crystal structures (from which information was extracted manually) to a remarkable 750,000+ crystal structures that can (in many cases) be interrogated automatically. Perhaps most importantly, many tools for extracting and evaluating this 'library' of information have been developed by the CCDC, allowing *non-expert* users to benefit from this invaluable source of experimental information. Interestingly, the CSD was described as the ‘fire’ which catalysed discussions and ideas over the last 50 years. This resonated strongly with me as, without the CSD, my own particular research area (leveraging prior structural information to improve the performance of crystal structure determination from powder diffraction data) would be even more challenging, denied of a source of valuable experimentally-derived structures that function as an ensemble.

The overall emphasis was that, as new challenges arise, better collaborations between the industrial and academic communities will be needed to overcome them. And with academic funding increasingly placing emphasis on the 'impact' that will result from the funding, it is hard to argue against this.

Etter MC (1990) ENCODING AND DECODING HYDROGEN-BOND PATTERNS OF ORGANIC-COMPOUNDS. Accounts of Chemical Research **23**:120-126.

Feeder N, Pidcock E, Reilly AM, Sadiq G, Doherty CL, Back KR, Meenan P and Docherty R (2015) The integration of solid-form informatics into solid-form selection. J. Pharm. Pharmacol. **67**:857-868.

Thomas LH, Wales C, Zhao L and Wilson CC (2011) Paracetamol Form II: An Elusive Polymorph through Facile Multicomponent Crystallization Routes. Crystal Growth & Design **11**:1450-1452.

The Next 50 Years

- Christer Aakeröy (Kansas State University) – "What are we going to do with all this information?"
- Colin Groom (CCDC) – "The CSD at 50: How will structural science look in another 50 years?"

Report by James McKenzie, University of Cambridge

Following an excellent display of how the CSD has been utilised past and present, the closing session of the CSD50 looked towards the future. **Christer Aakeröy** opened the session with his talk titled "What are we going to do with all this information?". The talk addressed some of the key issues that may arise in the future of an ever expanding CSD. The interesting question of "how much data is enough?" was raised. Does the CSD already hold enough data to answer all of our questions? Christer argued that we are still extremely far away from this point as there are still many classes of molecule or functional groups for which there is little data available. A comment from the audience furthered this point, adding that the content of the database is not uniformly distributed throughout the chemical space and we need crystallographers to fill in the missing gaps. The growth of the CSD will therefore remain vital for the continued elucidation of behavioural patterns which help guide industry and academia to making correct decisions.

"*You can't just put data out there*" – Olga Kennard

With the growth of the CSD comes a larger need for data curation. Christer acknowledged the fantastic work the CCDC do in scrutinising published structures to ensure their quality. This can only be achieved by a team of dedicated experts who are in constant contact with authors. Encouragingly, complementary technologies to crystallography may arise in the future, such as crystal structure prediction and microscopy. Christer argued that in order to consolidate all this information it will become even more important for the CCDC to curate, organise and distribute these data to the public.

The second and final talk was by **Colin Groom** who gave his predicted forecast of the changes that are likely to occur in the CSD, and how the CCDC will adapt accordingly. Colin predicted that the way in which that database is accessed will become easier, a point which was emphasised throughout the conference by numerous people viewing structures on their mobile phones and tablet devices. The types of structures that are deposited are likely to change with an increasing number of MOFs and nanoparticles being crystallised. Due to the increased size and complexity of these structures new graphics will be required to visualise them and new programs will be needed to analyse them. Additionally the number of deposits of host guest systems will increase due to the growing use of Fujita's crystal sponge method.
Colin explained how alternative funding models are being explored to ensure that the CSD is as far reaching and accessible as possible. The importance of overseas use

of the database was highlighted by Colin's prediction that in the near future the largest contributors to the CSD are likely to be researchers in China and India. An earlier talk by **Professor Zéphirin *Yav*** *from the University of Kinshasa* (D. R. Congo) explained how through collaboration with the CCDC they were able to access the CSD for not only teaching purposes but also to perform innovative scientific research. Future collaborations like this will be required to give access to the CSD and train people who would be otherwise unable to use it.