Institution: University of Reading		
Unit of Assessment: 3 (Allied Health Professions, Dentistry, Nursing and Pharmacy)		
Title of case study: DASH 3.3.7 solves challenging crystal structures of industrial importance.		
Period when the underpinning research was undertaken: Between October 2012 and October 2017		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Kenneth Shankland	Professor	Between 2009 and present
Period when the claimed impact occurred: Between November 2017 and present		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact

University of Reading research has enhanced, by an order of magnitude, the power of software to determine crystal structures from powder X-ray diffraction (PXRD) data. This has redefined the limits of structure determination for many important materials, including pharmaceuticals. The jump in performance achieved by the latest version of the software, DASH 3.3.7, allows for faster determination of more complex structures, providing vital solid-state structural knowledge to industrial users – for practical, reputational, and intellectual property purposes. Faster results are especially important to decision making during drug candidate selection in the pharmaceutical industry. The significant updates embedded in version 3.3.7 of DASH – already the leading computer program used to solve structures from PXRD data – expands its applicability to many more compounds. This provides a structural basis for understanding the behaviour of materials during processing. For example, this crystal structure knowledge enabled AstraZeneca to understand how the amorphous form of the company's irritable bowel syndrome (IBS) drug tenapanor (estimated peak sales up to USD 782,000,000, calculated December 2019) was stabilised during production.

2. Underpinning research

Structural information provided by X-ray crystallography is essential to fully understand the behaviour of materials during processing and use. As a result, crystal structure determination is used routinely by the pharmaceutical industry as well as the global food additives, agrichemical and dyestuff industries. If a single crystal of a substance cannot be grown for analysis, then powder X-ray diffraction (PXRD) can be used. Developed from concepts published in 1998 by Shankland and co-workers and sold by the Cambridge Crystallographic Data Centre (CCDC), DASH was already an important tool for structure determination by PXRD that had evolved steadily to improve useability and functionality, driven by Shankland's research at the University of Reading [1,2]. However, its core structure-solving algorithm remained unchanged.

In the past decade, industry users, including those involved in drug discovery programmes, have needed to obtain structural insights into molecules more complex than could normally be solved by DASH, and with faster turnaround times. To address this need, Shankland undertook fundamental research exploring how DASH's performance with complex structures could be improved by:

1. Optimising three key parameters controlling the core simulated annealing (SA) algorithm; and
2. Incorporating additional conformational information from known crystal structures in the Cambridge Structural Database (CSD).

Firstly, Shankland and his team investigated the effect of varying the three key SA control parameter values governing the position, orientation and conformation of a molecule within a crystal unit cell, using an advanced iterated racing approach. Shankland *et al*. performed optimization against a training set of 40 PXRD datasets. As a result, they were able to compare success rates in solving 100 crystal structures from PXRD datasets for both the established

default SA parameter values in DASH and the optimised SA parameter set suggested by iterated racing [3].

Secondly, the researchers optimised the ability of DASH to leverage **known** molecular conformation information derived from the CSD to improve program efficacy in solving **unknown** crystal structures. Using this novel approach, the SA sampling of rotatable torsion angles in the molecule being studied is biased by torsion angle distributions extracted from related structures in the CSD. The team then investigated the effect of introducing such biased sampling using 51 complex crystal structures selected from the aforementioned 100 PXRD datasets [4].

The outcome was that the best performing SA parameter set [3] and the biased sampling approach [4] each individually yielded significantly better performance (i.e. faster and more reliable crystal structure determination) across a wide range of structural complexities. **Used together, the researchers achieved an overall improvement in DASH performance of approximately one order of magnitude. This means that the probability of being able to solve such complex structures has been significantly improved by this research, and also that complex structures can be solved at industrially and academically relevant timescales.**

These findings led directly to an important update package, released as version DASH 3.3.7 in November 2017.

3. References to the research

The research outlined in this case study resulted from a long-term collaboration with the CCDC, which recognises the ongoing relevance of DASH to both the CCDC and its users in both academia and industry. Key findings from this partnership have been published in leading international journals and, in particular, as references [3] and [4] detail, the work has enabled DASH to remain as the world-leading tool for solving structures from PXRD data. Such structures also add collective value to the CSD, the definitive database of molecular crystal structures.

1. Spillman, M. J, **Shankland, K**., **Williams, A. C.,** and Cole, J.C. (2015). 'CDASH: a cloud-enabled program for structure solution from powder diffraction data'. *J. Appl. Cryst*. **48**, 2033-2039. DOI: https://doi.org/10.1107/S160057671502049X
2. **Kabova, E. A.,** Blundell, C. D. and **Shankland, K.** (2018). 'Pushing the Limits of Molecular Crystal Structure Determination From Powder Diffraction Data in High-Throughput Chemical Environments'. *J. Pharm. Sci*. **107** (8), 2042-2047. DOI: https://doi.org/10.1016/j.xphs.2018.04.010
3. **Kabova, E. A**., Cole, J. C., Korb, O., Lopez-Ibanez, M., **Williams, A. C.** and **Shankland, K.** (2017). 'Improved performance of crystal structure solution from powder diffraction data through parameter tuning of a simulated annealing algorithm'. *J. Appl. Cryst*. **50**, 1411-1420. DOI: https://doi.org/10.1107/S1600576717012602
4. **Kabova, E. A.**, Cole, J. C., Korb, O., **Williams, A. C**. and **Shankland, K.** (2017). 'Improved crystal structure solution from powder diffraction data by the use of conformational information'. *J. Appl. Cryst.* **50**, 1421-1427. DOI: https:/doi.org/10.1107/S1600576717012596

4. Details of the impact

The DASH 3.3.7 release – as distinct from other more routine DASH updates – represented a step-change in the performance of the leading computer program for solving crystal structures from PXRD data. This was vital to meet the evolving needs of industrial users, especially in the pharmaceutical industry. Since 2017, all users have benefitted from this enhanced ability to solve a significantly larger proportion of far more complex structures, faster than was previously possible [E1].

Prior art in powder diffraction

Sold by the CCDC since the year 2000, DASH is well-known to the pharmaceutical sector as *"the most successful software package* [derived] *from* [structure solution from PXRD] *research*" [E2]. The utility of DASH, beyond the pharmaceutical industry, was notably demonstrated in 2012 when it was used to solve three previously intractable crystal forms of Pigment Red 57:1, the most widely used red dye in printing (annual sales of EUR 200,000,000, calculated in May 2012) [E3].

Meeting the needs of an evolving drug discovery sector

Sector-wide interest in developing new targeted treatments for chronic diseases means that small-molecule drugs are being synthesised faster than ever before. The small-molecule drug market was valued at USD 25,000,000,000 in 2019 – a figure set to double by 2027 (calculated October 2020). Despite their name, small-molecule drugs are (on average) increasing in size and molecular complexity. It is, therefore, vital to the pharmaceutical industry that structure-solving techniques designed to characterise these drugs keep pace with these developments. The larger the number of crystalline phases of a drug that can be characterised, the better the relationships between them can be understood. This is a key step in moving a particular solid form of a drug towards marketing approval, and in avoiding intractable solubility and bioavailability limitations.

To meet this changing industrial need, and as part of a long-term research collaboration, the University of Reading and the CCDC invested in a PhD-based project to improve the ability of DASH. **The significant benefits offered by Shankland's two key findings [3, 4], enabled the unequivocal recommendation to CCDC's management that these new approaches should be used in a production environment.**

As a direct consequence, and following additional investment by CCDC, the University of Reading's optimised SA parameters became the new default setting in a major update, DASH 3.3.7, released in November 2017. This release also promoted the use of biased sampling. A total of GBP 125,000 (GBP 90,000 CCDC, GBP 35,000 University of Reading) has been invested in these two developments over this REF period, keeping DASH at the leading edge of structure solution capability [E1].

Reach and significance

DASH 3.3.7 allows larger crystal structures to be solved faster and more reliably than ever before, enabling researchers in industry and academia to develop a more complete understanding of commercially important materials in the solid state. Since November 2017, everyone using DASH has solved PXRD data using the 3.3.7 updates, and they are now able to solve more data sets, at a faster pace. This was independently confirmed by DASH 3.3.7 users solving previously intractable structures [E4]**.** Sold as part of the CCDC's Cambridge Structural Database System (CSD), DASH 3.3.7 reaches 1,166 academic sites worldwide in 87 countries. Industrial users of the CSD have the option to purchase DASH as part of the add-on 'CSD-Materials' module, and 21 of its 77 industrial users (i.e. those with a direct interest in the solid-state behaviour of small molecules and the in-house analytical expertise to interpret PXRD data) have chosen to take this option [E1].

The value of DASH 3.3.7 to pharmaceutical innovators

Innovators particularly value the enhanced structure-solving power of DASH 3.3.7, using it not only to further drug development, but also to provide vital solid-state information for patent protection and litigation, to satisfy regulatory authorities and to predict and rationalise important physicochemical properties, such as solubility [E5]. An outstanding example is the use of DASH in 2018 by AstraZeneca (market capitalization value USD 142,000,000,000) to solve the crystal structure of its development compound tenapanor dihydrochloride [E6, Figure 1]. Comprising 144 atoms, tenapanor dihydrochloride was an extremely challenging crystal structure to solve using PXRD.

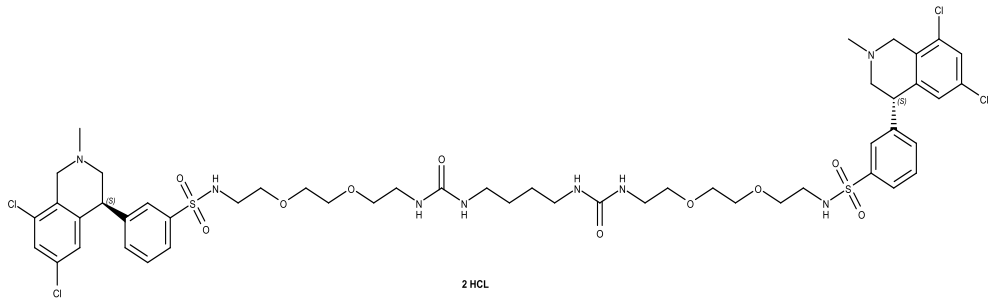

Figure 1: Tenapanor dihydrochloride: an example of the large size and high conformational flexibility of molecules now routinely solvable using DASH.

Tenapanor was approved by the US Food and Drug Administration in September 2019 for use in the treatment of irritable bowel syndrome with constipation, and is estimated to achieve a value of USD 200,000,000 in seven major markets by 2023. The knowledge of tenapanor's crystal structure provided through DASH 3.3.7 was vital to understand how the marketed amorphous form was stabilised during production. This level of solid-form understanding is key to allowing companies to move a particular solid form towards market. Without it, companies risk the therapeutic, financial and reputational consequences of the kind typified by the costly, high-profile product recall of ritonavir by Abbott Laboratories in 1998, after it was discovered that a less soluble crystalline phase of ritonavir appeared in the formulation over time, drastically changing its dissolution characteristics and rendering the drug ineffective.

The value of DASH to generic drug companies

Novel crystal structures of established drugs are of particular interest to generic pharmaceutical manufacturing companies, who look to develop and characterise them both for their own production/intellectual property considerations and for possible use in circumventing innovator patents. For example, Teva (market capitalisation value USD 10,500,000,000, November 2020) have used DASH to solve the form B-structure of ribociclib succinate, a drug with sales of USD 235,000,000 in 2018 [page 28, E7].

Redefining the limits of complexity

Structural complexity is often characterised by the number of parameters that need to be determined by DASH in order to solve a crystal structure. Before DASH 3.3.7, it was generally accepted that solving problems with >20 parameters was challenging. Now, problems with >30 parameters are being solved 'routinely' [E4] and previously intractable structures such as tenapanor hydrochloride (which has 45 parameters) are now tractable. This step change in capability, made possible by Shankland and his team's DASH 3.3.7 developments, has thus redefined the limits of PXRD-based structure determination, enabling the characterisation of many important, often high-value crystal structures that would be otherwise inaccessible. An extensive literature search revealed 26 compounds in 19 therapeutic areas, which have been solved using DASH since January 2018 [E8].

In addition, such crystal structures have (whether determined by industry or academia) not only specific value in the context of the material, but also collective value; most of them are deposited in the CSD, helping to fully populate the crystal structure landscapes of these commercially important materials.

As evidenced in user questionnaires [E5], DASH 3.3.7 is currently being used by some of the biggest innovator pharmaceutical companies in the world (including GlaxoSmithKline plc, AstraZeneca and Eli Lily) and some of the biggest generic companies (such Teva [E7]) in the high priority area of solving crystal structures that would otherwise be inaccessible due to a lack of single crystals. Used at many points in the drug development workflow, the crystal structures DASH 3.3.7 delivers are of high importance to the developability and patent protection of small molecules, which make up approximately 75% of new drug approvals. Companies view DASH as the world-leading software package of its type and the new capability in DASH 3.3.7 is viewed as of particularly high importance as it allows them to study more complex compounds (the current trend in small-molecule development) and solve their

structures faster, thus enabling more agile decision-making for onward drug development. Structures solved using DASH 3.3.7 **increased the number of new drugs reaching patients** and **saved substantial wasted development effort** by allowing the early identification of sub-optimal structures. It is also being applied to the structures of compounds currently on their way to being marketed as new drugs [E5].

5. Sources to corroborate the impact

[E1] Email correspondence with CCDC about DASH usage

[E2] Diamond Light Source synchrotron Phase III Beamline Proposal 058-S Large Area Detector Powder Diffraction Beamline, March 2011

[E3] Bekö S. L., Hammer S. M., and Schmidt M. U. (2012). 'Crystal Structures of the Hydration States of Pigment Red 57:1'. *Angew. Chem. Int. Ed.* **51**, 4735-4738. DOI: https://doi.org/10.1002/anie.201109082

[E4] Schlesinger, C., Bolte, M. and Schmidt, M.U. (2018). 'Challenging structure determination from powder diffraction data: two pharmaceutical salts and one cocrystal with Z′ 2'. *Z. Krist.* **234** (4), 257-268. DOI: https://doi.org/10.1515/zkri-2018-2093

[E5] Questionnaires completed by representatives of AstraZeneca, GlaxoSmithKline plc and Eli Lily on the value of DASH to their organisation

[E6] Nilsson Lill, S. O., Widdifield, C. M., Pettersen, A., Ankarberg A. S., Lindkvist, M., Aldred, P., Gracin, S., Shankland, N., Shankland, K., Schantz, S and Emsley, L. (2018). 'Elucidating an Amorphous Form Stabilization Mechanism for Tenapanor Hydrochloride: Crystal Structure Analysis Using X-ray Diffraction, NMR Crystallography, and Molecular Modeling'. *Mol. Pharmaceutics*. **15** (4), 1476–1487. DOI: https://doi.org/10.1021/acs.molpharmaceut.7b01047

[E7] International patent application (WO2019/040567): 'Ribociclib salts and solid state forms thereof'.

[E8] Table of drug structures solved using DASH and published between January 2018 and June 2020