

<b>Institution:</b> University of Cambridge
<b>Unit of Assessment:</b> UoA1
<b>Title of case study:</b> Accelerating structural biology with Phaser crystallographic software
<b>1. Summary of the impact</b> (indicative maximum 100 words) Knowledge of the three-dimensional structures of macromolecules is a prerequisite for understanding their function at the atomic level, an essential component of modern drug development. Most structures are determined by X-ray crystallography: the majority using molecular replacement (MR, which exploits known structures of related proteins), and about half of the remainder using single-wavelength anomalous diffraction (SAD). The Phaser crystallographic software, developed by Read and colleagues, implements powerful new likelihood-based methods for MR and SAD phasing and has made a large impact, accelerating over the period 2008-2013. At the pharma giant, AstraZeneca, Phaser is considered the "tool of choice" for solving structures by MR.
<b>2. Underpinning research</b> (indicative maximum 500 words) The conception and development of Phaser (and its predecessor Beast) have all taken place since Read became Professor of Protein Crystallography, on his arrival as a Wellcome Trust Principal Research Fellow at the Department of Haematology, University of Cambridge in 1998, though it continues a long-running theme of research into the application of likelihood to crystallography. The research towards the development of Phaser was led by Read and conducted by a team of post-doctoral researchers based in the Department of Haematology in the Cambridge Institute for Medical Research: Airlie McCoy (2000-present), Laurent Storoni (2001-2004), Hamsapriye (2004-2006), Gábor Bunkóczi (2007-present) and Robert Oeffner (2007-present). Traditional methods for solving protein crystal structures by molecular replacement (MR) suffer from a number of drawbacks, largely arising from the inability of these methods to take account of the effects of errors such as differences between the known structure and the unknown target. Maximum likelihood provides a way to account statistically for such errors; likelihood targets for MR searches were derived by Read and implemented in the computer program Beast, and were indeed shown to be significantly more sensitive (Read, 2001). Beast was very slow, but success in determining several difficult unsolved structures encouraged the development of a faster, more powerful new program, Phaser. Speed was increased by deriving and implementing fast approximations to the likelihood targets for orientation (Storoni <i>et al.</i> , 2004) and translation searches (McCoy <i>et al.</i> , 2005). Automation algorithms, built on the advantages of likelihood for decision-making, made it much easier to solve the structures of large complexes at the forefront of structural biology in both academia and industry. The first version of Phaser was released to the crystallographic community in late 2003, through open-source downloads to academic users, and to industrial users as part of the CCP4 and Phenix packages. The development of a likelihood target for the SAD phasing experiment (McCoy <i>et al.</i> , 2004) next gave Phaser the power to solve novel structures with no prior structural knowledge. Facilitated by the unified underlying mathematical foundation of the MR and SAD likelihood targets, combined methods were developed and implemented, allowing different sources of information to be used together in solving particularly recalcitrant structures (McCoy <i>et al.</i> , 2007; Read and McCoy, 2011). Phaser is still under continuous development in the Read lab to improve the algorithms and automation features. New versions are released formally, as part of the CCP4 and Phenix crystallographic software packages, about twice each year. The increased sensitivity of the likelihood targets in Phaser, compared to methods used previously, has opened new applications of the molecular replacement method. Read has collaborated to combine Phaser with the advanced modelling techniques of the Rosetta program (David Baker, the University of Washington), making it possible to solve crystal structures using <i>ab initio</i> folding models (Qian <i>et al.</i> , 2007); with the further addition of automated rebuilding software (contributed by Tom Terwilliger, Los Alamos National Laboratory), structures could be solved with considerably more distantly-related starting models than previously possible (DiMaio <i>et al.</i> , 2011). Read is also collaborating with Isabel Usón (Molecular Biology Institute of Barcelona) to strengthen her Arcimboldo procedure for <i>ab initio</i> structure solution, which uses Phaser to place small molecular

## Impact case study (REF3b)

fragments such as helices that seed completion of the rest of the structure.

**3. References to the research** (indicative maximum of six references)

- Read, R.J. Pushing the boundaries of molecular replacement with maximum likelihood. 2001. *Acta Cryst. D57*: 1373-1382. PMID: 11567148. Citations: 548. Journal impact factor: 14.1
- Storoni, L.C., McCoy, A.J. and Read, R.J. Likelihood-enhanced fast rotation functions. 2004. *Acta Cryst. D60*: 432-438. PMID: 14993666. Citations: 834. Journal impact factor: 14.1
- McCoy, A.J., Storoni, L.C. and Read, R.J. Simple algorithm for a maximum-likelihood SAD function. 2004. *Acta Cryst. D60*: 1220-1228. PMID: 15213383. Citations: 39. Journal impact factor: 14.1
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. Likelihood-enhanced fast translation functions. 2005. *Acta Cryst. D61*: 458-464. PMID: 15805601. Citations: 1177. Journal impact factor: 14.1
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. Phaser crystallographic software. 2007. *J. Appl. Cryst.* 40: 658-674. PMID: 19461840. Citations: 2968. Journal impact factor: 3.3
- Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J. and Baker, D. High-resolution structure prediction and the crystallographic phase problem. 2007. *Nature* 450: 259-264. PMID: 17934447. Citations: 145. Journal impact factor: 38.6
- DiMaio, F., Terwilliger, T.C., Read, R.J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H.L., Das, D., Vorobiev, S.M., Iwai, H., Pokkuluri, P.R. and Baker, D. Improved molecular replacement by density- and energy-guided protein structure optimization. 2011. *Nature* 473: 540-543. PMID: 21532589. Citations: 43. Journal impact factor: 38.6
- Read RJ and McCoy AJ. Using SAD data in Phaser. 2011. *Acta Cryst. D67*: 338-344. PMID: 21460452. Citations: 12. Journal impact factor: 14.1.

**4. Details of the impact** (indicative maximum 750 words)

X-ray crystallography has grown to be one of the pillars of research and development in the pharmaceutical industry, where the atomic interactions with drug targets of many candidate molecules are studied to optimise them in the development of new drugs. Industrial crystallographers examine numerous complexes with the same target or many examples from "druggable" families of proteins such as kinases, so molecular replacement is one of their most important tools.

Phaser met a real need for both academic and industrial crystallographers, indicated by its rapid adoption in preference to previous programs for carrying out molecular replacement calculations. After it was released in 2003, it quickly caught on because of success in solving a number of structures that had resisted years of effort. The 1000<sup>th</sup> download was marked within 15 months of the initial release. Already by 2008, 1190 of 6248 X-ray crystal structures (19%) released in the Protein Data Bank (PDB: [www.rcsb.org](http://www.rcsb.org)) cited the use of Phaser (ref. 1). From 2008, this has continued to grow: 1648 of 6746 (24%) structures released in 2009, 2122 of 7296 (29%) in 2010, 2705 of 7468 (36%) in 2011, 3067 of 8302 (37%) in 2012 and 2367 of 5962 (40%) up to the end of August in 2013. Given that X-ray crystal structures account for about 90% of new entries in the PDB, Phaser has accounted for over 1/3 of all new macromolecular structures in the last three years.

Though most structures in the PDB are contributed by academic researchers, it should be noted that the pharmaceutical industry makes heavy use of these data, including the many structures solved with the use of Phaser. Industrial scientists have also rapidly adopted Phaser, for the same reasons as their academic colleagues.

Specific examples of impact in the pharmaceutical industry are documented in two letters. A research fellow at Bristol-Myers Squibb (ref. 2) describes several cases in which the use of Phaser allowed the solution of structures that had previously been difficult or even impossible. In one specific example he describes working on the structure of a biologic/target complex, where he had only a limited amount of protein and a limited number of crystals and for which he states that Phaser was "crucial to the determination of this structure". An Associate Principal Scientist at AstraZeneca (ref. 3) states that "Phaser has been instrumental in solving several target structures recently, and helped the progress of these projects by making a costly and lengthy experimental phasing unnecessary, which would otherwise be a bottleneck in a structure based drug discovery campaign". AstraZeneca employs about 30 FTEs in structural biology, and they "consider Phaser

## Impact case study (REF3b)

as a tool of choice when solving novel structures by molecular replacement". She also states that "Phaser outperforms other programs and gives better confidence in the solution". Both of these researchers in industry emphasise that, by making difficult problems easy, valuable time is saved.

We have clear evidence of wider take-up by industrial users. Licences to use Phaser are available as part of two packages: CCP4 (about 120 site licences of the package, at \$9500 per licence, to industry including AstraZeneca, Bristol-Myers Squibb, GlaxoSmithKline, Hoffmann-La Roche, Merck, Novartis and Vertex Pharmaceuticals, ref. 4) and Phenix (13 industrial participants in its consortium, ref. 5). A search of US patents (ref. 6) reveals that 42 patents filed since the beginning of 2008 cite the use of Phaser in the research underlying the new intellectual property. Considering that there is an average of nearly three years between these patent applications being filed and granted, this is very much a lower bounds estimate of the impact of Phaser on the development of new IP. These patents have been assigned to a variety of entities, including Genentech, Janssen Pharmaceutica, Novo Nordisk and, in the UK, MedImmune and Heptares Therapeutics.

The Phaser development team has answered queries about the use of Phaser from scientists at 22 different companies, including Abbott, AstraZeneca, Bristol-Myers Squibb, Johnson&Johnson, Heptares Therapeutics, Novartis and Sanofi-Aventis. In addition, industrial crystallographers attend the annual CCP4 Study Weekend and the biannual Phenix Developers' Workshop, where they take the opportunity to ask questions about the use of Phaser and to request new features.

Industrial royalty revenues received by the Phenix team are shared among the partners. The Cambridge share of about £180,000 to March 2013 has been distributed among the University, CIMR, Catalyst (Wellcome Trust) and the Phaser developers.

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

1. Statistics on Phaser usage were obtained from the PDB search facility: [www.rcsb.org/pdb/search/advSearch.do](http://www.rcsb.org/pdb/search/advSearch.do), searching for "phaser" in the "Text Search" query type.
2. Letter from Research Fellow, Protein Science and Structure, Bristol-Myers Squibb Research and Development, 1 January 2013.
3. Letter from Associate Principal Scientist, Structure and Biophysics, Discovery Sciences, Astra Zeneca. 23 January 2013.
4. CCP4 industrial licence holders are listed each year in the special edition of Acta Crystallographica Section D containing the proceedings of the annual CCP4 Study Weekend, published most recently in part 4 of volume 68, April 2012.
5. Phenix industrial consortium members are listed at <http://www.phenix-online.org/consortium/participants/>
6. The US PTO website was searched by looking for granted patents containing the terms "Phaser" and "crystal", and filed since the beginning of 2008, by using the query "APD/1/1/2008->12/31/2013 and phaser and crystal" in the advanced search tool at <http://patft.uspto.gov/netahtml/PTO/search-adv.htm>, then verifying whether the Phaser technology was indeed referenced in each patent.