

Impact case study (REF3b)

Institution: University of Oxford
Unit of Assessment: 10 – Mathematical Sciences
Title of case study: Exploitation of rapid protein structure prediction tools
1. Summary of the impact <p>Novel rapid methods for predicting protein structure, particularly functional loop structures, have been developed by researchers at the University of Oxford. These have been made accessible to a large audience through a suite of computational tools. The methods have had general impact through download and online access and specific impact through extensive use within UCB Pharma. The tools are much faster than other methods, creating equal or better predictions in approximately a thousandth of the time. Commonly exploited by UCB Pharma in their drug discovery pipeline, they have cut computational cost, but, more importantly, they have greatly reduced the time for process improvements. UCB Pharma estimate that the tool pyFREAD alone saves over £5 million in the discovery costs for a single drug molecule. FREAD (a version of pyFREAD coded in C) is also being used more widely, for example by Crysalin Ltd and InhibOx.</p>
2. Underpinning research <p>Proteins perform crucial functions in all biological processes and in general their function is specified by their three-dimensional structure. Understanding a protein's structure is an essential step in the drug discovery process and most pharmaceutical companies make extensive use of structural information. Experimental methods can provide some information but are not always applicable or able to give complete structural information. One of the areas where experimental methods most commonly fail is in the loop-structures of proteins. It is these loops which tend to determine the function of proteins. Thus computational methods for loop prediction offer a powerful addition to the experimentally available data.</p> <p>Loop prediction has been approached in two ways, ab initio and database search. In recent years it had been thought that ab initio methods were more powerful. Researchers at the University of Oxford, led by Professor Charlotte Deane, identified that database methods had been underestimated and found that sequence similarity, as quantified by environment-specific substitution scores, can be used to significantly improve prediction [1]. The research gives a method for predicting loop-structures and for calculating any missing structural data.</p> <p>In 2010, Deane and her team at the University of Oxford rewrote the computer program FREAD (and wrote pyFREAD) to incorporate a completely new scoring system which, combined with bigger databases of protein structures and faster computers, resulted in a significant improvement in the ability to predict the protein structure. These improvements in prediction and speed are reported in [1], see for example Figure 1.</p> <p>pyFREAD has been demonstrated to show higher accuracy than comparable tools such as the loop refinement modules in the commercially available Prime or MODELLER packages and to be significantly faster than its competitors. Professor Deane and collaborators at the company UCB Pharma realised that the pyFREAD methodology was also applicable to the much more general problem of model completion and, given its speed and accuracy, it could also be used when multiple segments of data were missing. A specific version of the method was then developed within pyFREAD in order to consider this more general problem and to address specific issues of interest to UCB Pharma. This second phase was completed in September 2012 and written up in [2], which includes predictions applied to membrane proteins. The significant speeding up of the algorithms has been shown, as well as the generalisation of the method to allow any fragment of the protein, not just loop structures, to be considered. Paper [2] also demonstrates that results of an even higher quality are obtained by using just the databases most appropriate to a given protein.</p>

Impact case study (REF3b)

The key researcher, Prof Charlotte Deane has been a University Lecturer since joining Oxford in 2002.



Figure 1. The predictive power of FREAD - An example prediction. The black structure is the correct answer structure. The grey loop is the top prediction without use of the environment specific substitution score. The white loop is the top prediction by FREAD using the environment specific substitution score. Reproduced from reference [1].

3. References to the research

- * [1] Choi Y, Deane CM, FREAD revisited: Accurate loop structure prediction using a database search algorithm, *Proteins*, 2010, 78(6), 1431-40. DOI: 10.1002/prot.22658.
- * [2] Sebastian Kelm, Anna Vangone, Yoonjoo Choi, Jean-Paul Ebejer, Jiye Shi, Charlotte M. Deane, Fragment-based modelling of membrane protein loops - successes, failures and prospects for the future, *Proteins*, 2013. DOI: 10.1002/prot.24299.

The two asterisked outputs best indicate the quality of the underpinning research. *Proteins* is an international refereed journal.

4. Details of the impact

The impact of the research falls into the category of economic benefit which we illustrate through the benefits realised by UCB Pharma. Other pharmaceutical companies such as Crystalin Ltd and InhibOx have also benefited. The research also has downstream impact on patient health.

Pathway to impact

pyFREAD has been made accessible through three different routes:

- the direct implementation in 2010 of the software by UCB Pharma, who were industrial partners in the research project. The Director of Computational Structural Biology at UCB Pharma states [A] "... *Professor Deane kindly provided to us at no charge the software pyFREAD, which was developed in her laboratory*"
- a web-based computational tool (<http://opig.stats.ox.ac.uk/sites/fread/>). The Head of Computational Chemistry, Crystalin Ltd, states in a letter in October 2013 [E] "*I downloaded it from your website ... last August*".
- a freely downloadable version of the FREAD software, published in 2010, (<http://opig.stats.ox.ac.uk/webapps/fread/php/>)

Nature and extent of the impact

The impact of pyFREAD is most readily measured through its use by UCB Pharma. They, as well as other major pharmaceutical companies, use X-ray crystallography and molecular dynamics simulations to guide 'lead optimization' in an iterative fashion. Atomic interactions between each compound and the target protein are analyzed and chemical modifications to the lead compound are designed accordingly to improve potency and selectivity. However, X-ray structures of the compound-protein complex often have undefined residues due to experimental limitations; such

Impact case study (REF3b)

residues must be modelled before dynamic simulations can be carried out.

The Director of Computational Structural Biology at UCB Pharma states [A] *“We used to rely on the software Prime from Schrodinger Inc to model the undefined residues in X-ray structures. [...] This software [pyFREAD] was at least 1000 times faster than Prime and also produced more accurate results! For example, to model a stretch of 14 undefined residues, it took 50 CPU hours with Prime but less than a minute with pyFREAD. In another occasion, we tried to reconstruct 7 stretches of undefined residues within the same X-ray structure; Prime crashed after running for 3 days without producing any useable results, while pyFREAD managed to generate accurate models in merely 3 minutes. Pleasantly surprised by the lightning speed and accuracy of this software, we immediately switched to using pyFREAD for such tasks and have not been disappointed.”*

He further summarises the immediate financial benefits [A] *“It is immediately clear that pyFREAD saves us not only £45,000 in annual license fee for Prime, but also thousands of CPU hours for each lead optimization campaign.”*

UCB Pharma operates in 40 countries worldwide and had a global revenue of €3.4 billion in 2012. They identify (in [A]) that lead optimization is one of the most costly steps in drug discovery and development, requiring on average £6 million per campaign (just one stage of the drug discovery process). In order to bring an approved drug onto the market, typically 15 lead optimization campaigns have to be carried out, costing a total of £90 million. The Director of Computational Structural Biology at UCB writes in [A] *“... that switching to pyFREAD shortened each lead optimization cycle by an average of 3 days. A typical lead optimization campaign lasts 2 years, the cost is £3 million per year, or £58,000 per week. Shortening each optimization cycle by 3 days translates to £35,000 in direct savings per iteration, or £350,000 for a 10-iteration campaign. As 15 campaigns are needed to put one drug onto the market [6], the total savings per drug approval achieved by using pyFREAD is expected to be over £5 million.”* This corresponds to saving UCB at least 5% for each programme in which it has been used.

The impact for UCB of this research goes beyond the financial savings. UCB state that [A] *“The research work at Professor Deane’s laboratory has generated significant economic value for UCB Pharma through the acceleration of the drug discovery process. More importantly, faster drug discovery means that patients receive better treatment sooner. While the impact on patients’ quality of life is hard to quantify, it is what matters most”.*

Evidence of wider, less quantifiable, impact of the FREAD methodology is seen from its web-based computational version. The data available for 2013 show this has performed on average over 60 predictions per month and was visited by over 200 unique users per month from throughout the world [B]. As exemplars of the reach of FREAD, the Chief Executive Officer of InhibOx, who have been using the software since 2011, states [C] *“I am writing to confirm that InhibOx has used the program FREAD to support our drug design work. The program was used primarily by one of our staff to assist in a project to build homology models of a malaria parasite serine protease, which we have been working on as a novel anti-malarial drug target”*, while the Head of Computational Chemistry at Crysalin Ltd says [D] *“Thank you for permitting me to use FREAD at Crysalin.... [it] is providing to be an invaluable addition to our existing tools.”*

5. Sources to corroborate the impact

[A] Letter from the Director of Computational Structural Biology UCB Pharma, describing their use of pyFREAD and the significance of the impact. Copy held by the University of Oxford.

[B] Data from webserver and software pages for pyFREAD:

<http://opiq.stats.ox.ac.uk/webapps/fread/php/>

<http://opiq.stats.ox.ac.uk/sites/fread/>

Copy held by University of Oxford

Impact case study (REF3b)

- [C] Letter from the Chief Executive Officer, InhibOx, describing their use of FREAD. Copy held by University of Oxford.
- [D] Letter from the Head of Computational Chemistry, Crysalin, describing their use of FREAD. Copy held by University of Oxford.
- [B], [C] and [D] provide examples of the reach of the impact of FREAD and pyFREAD.