Impact case study (REF3)

Institution: Imperial College London

Unit of Assessment: 5 – Biological Sciences

Title of case study: A5-2 Commercial impact of Phyre: A resource for computational modelling of the 3D structure of proteins

Period when the underpinning research was undertaken: 1 Oct 2001 – 31 Dec 2020

Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Prof Michael Sternberg	**Role(s) (e.g. job title):** Director Centre for Integrative Systems Biology and Bioinformatics	**Period(s) employed by submitting HEI:** 2001 – to date

Period when the claimed impact occurred: 1 Aug 2013 – 31 July 2020

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact (indicative maximum 100 words)

Phyre is an internationally-used software suite with a powerful web interface from Sternberg's group at Imperial College London for predicting protein 3D structure from sequence. Non-academic users include pharma, biotech and agriculturally-focussed companies, and these range from international organisation to SMEs. Commercially and societally impact include enhanced understanding of the structural consequences of genetic mutations associated with disease, the design of vaccines and therapeutic antibodies, functionally annotating food and pathogen genomes, and providing structural models for drug discovery. Since 1 August 2013, there were over 500,000 distinct users from over 80 countries including over 900 commercial users. Over 70 granted patents cite Phyre.

2. Underpinning research (indicative maximum 500 words)

Knowledge of the 3D structure of proteins is central to understanding its function and guides studies including the explanation of the molecular basis of disease, the development of modified proteins and novel therapeutics, the structural and functional annotation of genomes, and the provision of structural targets for computer-aided drug discovery. There is an ever-expanding gap between the numbers of determined protein sequences and structures; today there are >150M sequences but only <100,000 different structures. Algorithms can be developed successfully to predict structure from sequence and the most accurate approach, which is used by Phyre, is modelling an unknown structure on the known structure of an evolutionarily-related protein (i.e. the template).

In 2001 Sternberg moved to Imperial where they hosted the first generation of their protein structure prediction resource (3D-PSSM). At Imperial, worked commenced in 2001 on developing a totally new package **Phyre1** (pronounced FIRE1). Phyre1 had a totally new core algorithm to recognise remotely-related sequences and their known structures based on a benchmarked consensus of multiple prediction techniques **[1,2]**. An entirely new interface was developed including interactive 3D molecular visualisation.

In 2011, **Phyre2 [3]**, was launched resulting from another complete re-write of the system now using hidden Markov models as the core search algorithm. The development of Phyre was guided by the user support we provided for Phyre1. The interface was totally redesigned to benefit from developments in web technology such as JavaScript, and Ajax. Several new backend components which provided extensive functionality were developed making Phyre2 unique in the world. These included:

Poing – a novel algorithm to combine models and to model stretches of proteins with no detectable template **[4]**.

SuSPect – a novel algorithm to predict the phenotypic effect of a missense variant, which is central to the interpretation of disease-associated genetic variants **[5]**.

3Dligand-site [5] – a novel algorithm to identify the location in a protein where a ligand is predicted to bind, which provides powerful insights to guide drug discovery **[6]**

A batch mode for high-throughput processing of proteomes (1,000s of sequences).

Phyre Investigator to examine models in detail for structural quality, functional sites, SuSPect predictions, 3D pockets, and interfaces.

One-to-one threading so users can build models based on their selected template.

PhyreAlarm to automatically inform the user if a superior predicted model becomes available.

Phyre1 and Phyre2 were always freely available to non-commercial users and in October 2018 software was made freely available to commercial users.

In 2019 the Sternberg group develop and launched (i) the **Missense3D** algorithm to evaluate the stereochemical impact of a missense genetic variant and (ii) the **PhyreRisk** database which enables human genetic variants to be mapped onto experimental and Phyre-predicted structures. PhyreRisk and Missense3D are freely available to the commercial community.

Phyre is now included as a "tool" in the European-centred network of key bioinformatics resources via the UK ELIXIR node and this further promotes its uptake by commercial and academic users.

The development of Phyre, Phyre2, Missense3D, PhyreRisk and was directed by Professor Sternberg whilst at Imperial (2001 – date). Prof Houlston (ICR) contributed to the development of PhyreRisk.

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

Citations from Google Scholar 27 Sept 2020
[1] Bennett-Lovsey, R. M., Herbert, A. D., **Sternberg, M. J.,** & Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: Structure, Function, and Bioinformatics*, *70*(3), 611-625. (99 citations) https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.21688

[2] Kelley, L. A., & **Sternberg, M. J**. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, *4*(3), 363. (4736 citations) https://www.nature.com/articles/nprot.2009.2

[3] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & **Sternberg, M. J.** (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858 (5032 citations) https://doi.org/10.1038/nprot.2015.053

[4] Jefferys, B. R., Kelley, L. A., & Sternberg, M. J. (2010). Protein folding requires crowd control in a simulated cell. *Journal of molecular biology*, *397*(5), 1329-1338. (99 citations) https://doi.org/10.1016/j.jmb.2010.01.074

[5] Yates, C. M., Filippis, I., Kelley, L. A., & **Sternberg, M. J.** (2014). SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of molecular biology*, *426*(14), 2692-2701 (139 citations) https://doi.org/10.1016/j.jmb.2014.04.026

[6] Wass, M. N., Kelley, L. A., & **Sternberg, M. J**. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic acids research*, *38*(suppl_2), W469-W473. (568 citations) https://doi.org/10.1093/nar/gkq406

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

Meeting the requirements of diverse users

The impact of software is generally dependent on both its power and its ease of use. Throughout the Phyre development cycle, we obtained extensive feedback from users about their specific needs and this directly drove the design of both the back-end algorithms and the web front-end. These mechanisms promoted extensive use by non-commercial users over many years and the reputation of Phyre led to its subsequent rapid and widespread uptake by commercial users.

Dissemination to non-academic users

In addition to academic publications, Phyre has been presented to both the academic and commercial communities at the tech track of the major bioinformatics meeting (ISMB) annually since 2015 and at over 12 UK training workshops including at Oxford and at the EBI **[A]**. Academic and commercial users can find out about Phyre via its listings on its own page on Wikipedia and sites listing bioinformatics software such as ELIXIR and ExPASy **[B, C]**. Additionally, we provide e-mail support. As a result of the quality of the algorithm, its user-friendly interface, its short turnaround time (typically a few hours for a single sequence) and our support, there have been a large number of academic publications reporting the use of Phyre. This has promoted Phyre to commercial users who consequently availed themselves of Phyre when it made available for their use.

Usage figures

In Phyre a user inputs a single protein sequence for analysis as this is recorded as a submission. Distinct users are based on a unique IP address. Over the period 1 Aug 2013 to 31 July 2020 there were over 500,000 distinct users who submitted over 1,000,000 jobs **[D]**. From 1 October 2018 to 31 July 2020 there were over 900 commercial users (as identified by clicking a radio button) and over 2,500 submitted jobs **[D]**.

Commercial impact

Since October 2018 Phyre2 has been available to commercial users, mainly in the pharmaceutical and drug discovery sectors. Here we give five specific company examples where Phyre2 has become an integral part of their discovery pipelines and has positively impacted their development programmes and businesses.

Evotec is a leading international Contract Research Organisation specialising in discovery, optimisation and development of small molecule and anti-body based therapeutics for a wide variety of pharmaceutical organisations. Phyre2 is now used to aid in construct and solving structures by predominantly crystallography but also cryo-electron microscopy. Evotec states that "*It* [Phyre2] *has contributed to >40 projects over the past year with a focus on establishing appropriate protein constructs to support crystallography and also high-throughput and associated biochemical and biophysical assays.*" Furthermore, this new approach has allowed them to "*identify dynamic loops and domain boundaries, which have been critical to the success of specific structural projects…*" **[E].**

Hummingbird Bioscience is a biotechnology company founded in 2015 based in Singapore and the USA. They develop novel bio-therapeutics to revolutionize treatment outcomes using a unique antibody discovery platform combined with detailed insights from systems and computational biology. Hummingbird Bioscience's Chief Technology Officer explains how Phyre2 is an integral part of their discovery platform. "*We utilize Phyre2 on a regular basis to inform our Rational Antibody Discovery platform to help model protein structures which lack experimentally determined structures and to place known structures into context. These are critical steps to ensuring that our molecules achieve a desired mechanism of action when binding a target molecule. Without Phyre2 we would not have been able to do this.*" **[F]**. This company is relativity young and currently has two molecules entering Phase I trials. Both targeted at indication in oncology. Phyre2 has been critical is both of these projects, in particular, in the selection of the targets and epitopes. Hummingbird Bioscience states that "*This would have been extremely difficult for the company at that early stage since commercially available tools were not financially*

viable at that stage. The fact that Phyre2 is freely available for commercial use was a crucial factor in the company's success." **[F]**

Vernalis is a structure-based drug discovery company of some 70 employees, based near Cambridge, UK. Its main activity is using our integrated suite of biophysical, structural, and medicinal chemistry methods in collaboration with international pharmaceutical companies in the discovery and optimisation of small molecules as modulators of the function of proteins of therapeutic relevance. Most of our target proteins are selected by the collaborator, so our use of bioinformatics tools is mainly limited to construction of models where crystal structures are not available, and analysis of the likely variation in function across species or arising from mutations. One user at Vernalis Research, writes "*it is our general experience that many of the methods available online have varying and sometimes obscure interfaces and the results generated are difficult to interpret. For us, the main advantage that Phyre2 brings is the intuitive interface to a collection of leading edge, robust methods, and the way the results are analysed and presented. The main impact on our research is that we can perform such analyses which inform the early stages of drug discovery projects.*" **[G]**

BASF is a multinational chemicals company with sales of €59 billion in 2019. Phyre has been used by the Belgium Innovation Centre, which has about 180 researchers and covers all aspects of the R&D process from the lab to the field. This Centre requires details of the 3D structure of proteins. A joint letter from the Site Director and Head of Intellectual Property states that "*Phyre has been successfully used in analysing a dozen proteins*". They describe the search algorithm as "*both accurate and efficient providing a turnaround time in a few hours which greatly facilitates our use in exploring hypotheses over the timescale required in our R&D pipeline*" **[H]**.

Domainex is a leading integrated drug discovery research service partner based near Cambridge, UK, offering tailored services, from expression of recombinant proteins, screening disease targets to identify hit compounds through to the discovery of pre-clinical drug candidates. The letter from a Principal Scientist at Domaniex states that "*Phyre is an extremely useful tool that quickly provides us with a large amount of target information. The structural prediction gives us information on potential domain boundaries, regions of disorder, potential ligand binding sites and the position of surface accessible lysine or cysteine residues that could be used for downstream labelling for assay development or biophysical analysis. The structural alignments provide us with direct links to related proteins that have already been successfully produced, providing us with established production methods.*" **[I]**. In particular they have integrated Phyre into Combinatorial Domain Hunting (CDH) technology to identify physical domain boundaries.

Academic use that translates into commercial and societal impact
Determination of protein structure and its implication for functional assignment is central to a wide range of academic research in bioscience and biomedicine that then translates into commercial and societal impact. A search of Google Patents with "PHYRE AND PROTEIN" identified over 70 granted patents that cite Phyre **[J]**. These patents cover a wide range of application areas including: vaccine design (e.g. US-10495639-B2) by Virginia Commonwealth University, anti-cancer kinase inhibitors (US-9198891-B2) by New York University, HIV envelope immunogenic polypeptides (US-9855329-B2) by Los Alamos & Dana-Faber & Duke, and staphylococcal antigens (EP-2488547-B1) by University of Edinburgh.

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

[A] A recent training course (Nov 2020) held at the EBI (European Institute of Bioinformatics) that includes Phyre – open to academia and industry
https://www.ebi.ac.uk/training/events/2020/structural-bioinformatics-virtual (Archived here)
[B] Phyre identified as a resource within ELIXIR UK https://elixiruknode.org/node-services/ (Archived here)
[C] Phyre identified as a resource on ExPASy https://www.expasy.org/ (Archived here)
[D] Print out of Google Analytics for access to Phyre (Archived here)
[E] Letter from Evotec

[F] Letter from Hummingbird Bioscience
[G] Letter from Vernalis
[H] Letter from BASF
[I] Letter from Domainex
[J] List of more than 70 granted patent identified via searching in Google Patents with "Phyre AND Protein" (Archived here)