Institution: University of Oxford

Unit of Assessment: 10: Mathematical Sciences

Title of case study: Computational Tools for the Discovery, Development and Design of Antibody Therapeutics

Period when the underpinning research was undertaken: 2014 – 2020

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:
Professor Charlotte Deane	Professor	Sept 2002 – present
Dr Konrad Krawczyk	Research Software Engineer	Feb 2016 – July 2018
Dr Claire Marks	Research Software Engineer	Nov 2018 – present
Dr Fergus Boyles	Research Software Engineer	Jan 2020 – present

Period when the claimed impact occurred: 1 August 2014 – 31 December 2020

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

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

Since 2014, researchers in the Oxford Protein Informatics Group (OPIG) at the University of Oxford have been developing a suite of freely available databases and open source predictive computational tools (collectively known as SAbDab-SAbPred and packaged together into SAbBox in 2020) to assist in the discovery and development of antibodies for use as therapeutics.

The suite has been used by over 100 international pharma companies, including GlaxoSmithKline (GSK) and UCB, as part of their antibody therapeutic development pipelines, both lowering the cost of antibody therapeutic development and increasing the speed at which such therapeutics can be delivered. This has also resulted in 71 patent applications in different parts of the world, all citing the use of SAbDab-SAbPred.

2. Underpinning research (indicative maximum 500 words)

Antibodies are essential proteins of the immune system that bind to potentially pathogenic molecules. Their ability to bind to a huge variety of substances with high specificity and affinity means that they are used as research tools and therapeutic agents. Since 2014, researchers in the Oxford Protein Informatics Group (OPIG) at the University of Oxford, led by Professor Charlotte Deane, have been building freely available databases and open source tools for the computational prediction of properties and characteristics of antibodies, such as binding affinity, stability and immunogenicity. The tools not only predict such properties but also suggest how the antibodies can be modified to improve therapeutic action. The underpinning research spans a range of topics including extensive data collection, software management, new algorithms for assigning structures, and development of vastly improved algorithms for rapid computation using known predictive models.

In 2014 the first database, **SAbDab** [1] the structural antibody database, was made freely available. SAbDab contains all publicly available antibody structures, and is updated automatically every week. Each structure is annotated consistently with properties including experimental details, antibody nomenclature, curated affinity data and sequence annotations. The database enables users to carry out many tasks including individual structure inspection, the creation and downloading of antibody sets with specific properties for analysis, finding structures with similar sequences to a query, and monitoring the overall known structural repertoire of antibodies.

The second key step in the development of the suite was **ANARCI** [2] which was released in 2015. This computational program numbers antibody sequences with any of the standard numbering schemes. It is the only downloadable, freely available tool for this purpose. Once an

antibody has been numbered it can be easily and accurately compared to all other antibodies, for example allowing users to identify if antibodies are closely related.

ABodyBuilder [3] is another of the key programs within the SAbDab-SAbPred suite, released in 2016. It generates three-dimensional models of an antibody from protein sequence information. Its accuracy is comparable to the best academic or commercial antibody modelling tools. Its unique features are that it annotates the three-dimensional model with an accuracy estimate, flags structural motifs that are known to cause issues during *in vitro* development of antibody therapeutics, and functions an order of magnitude faster than other methods, making it more useful in a therapeutic antibody development context where thousands of models are often required.

In 2016, to improve ease-of-use, the tools SAbDab, ANARCI and ABodyBuilder were unified under the name **SAbPred** [4] and provided through a web-based server that makes predictions of the properties of antibodies using the three-dimensional structure as a key step. Within SAbPred a user can number and align sequences and generate structural models of antibodies automatically. Furthermore, users can annotate such models with information including estimated accuracy, and sequence and structural properties. It also contains software to predict the epitope (the binding site on the protein).

One additional program added to SAbPred in 2019 was **TAP** [5] – the Therapeutic Antibody Profiler. To be successful therapeutics, antibodies need to be formulated so that they not only bind to their target of interest but are also “developable” (e.g. they are not immunogenic, unstable or self-associating). Developability issues often halt the progress of a potential antibody therapeutic. TAP is a rapid computational tool that can flag possible issues with a therapeutic antibody sequence. It highlights any characteristics that are rare/unseen in clinical-stage antibody therapeutics. As TAP builds on ABodyBuilder, it is able to flag these issues when only an antibody sequence is known.

A further addition to the suite in 2018 was **OAS**: Observed Antibody Space [6]. This is a semi-automated database that collects and annotates immune repertoires for use in large-scale analysis. Immune repertoires are experimental datasets of antibody sequences from an individual (human or animal). OAS contains over 5,000,000,000 sequences from 80+ studies. OAS sorts, cleans, annotates, translates and numbers all of the antibody sequences contained in the immune repertoires. This means users can filter the sequences based on attributes such as chain type, species and disease state, allowing identification of disease-critical properties.

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

All articles in internationally refereed journals.

[1] Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J. & Deane, C.M. (2014) SAbDab: the structural antibody database. *Nucleic Acids Res.*, 42 (Database issue): D1140-D1146, DOI: 10.1093/nar/gkt1043

[2] Dunbar, J. & Deane, C.M. (2015) ANARCI: Antigen receptor numbering and receptor classification. *Bioinformatics*, 32:298-300, DOI: 10.1093/bioinformatics/btv552

[3] Leem, J., Dunbar, J., Georges, G., Shi, J. & Deane, C.M. (2016) ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. *mAbs*, 8:1259-1268 DOI: 10.1080/19420862.2016.1205773

[4] Dunbar, J., Krawczyk, K., Leem, J., Marks, C., Nowak, J., Regep, C., Georges, G., Kelm, S., Popovic, B. & Deane, C.M. (2016) SAbPred: a structure-based antibody prediction server. *Nucleic Acids Res.*, 44: W474-W478, DOI: 10.1093/nar/gkw361

[5] Raybould, M.I.J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A.P., Bujotzek, A., Shi, J. & Deane, C.M. (2019) Five Computational Developability Guidelines for Therapeutic Antibody Profiling. *PNAS*, 116:4025-4030, DOI: 10.1073/pnas.1810576116

[6] Kovaltsuk, A., Leem, J., Kelm, S., Snowden, J., Deane, C.M. & Krawczyk, K. (2018) Observed Antibody Space: A resource for data mining next-generation sequencing of antibody repertoires. *J. Immunol.*, 201. (8): 2502-2509, DOI: 10.4049/jimmunol.1800708

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

Antibodies are the most important and successful class of bio-therapeutics with over 100 approved for clinical use in the US, and hundreds more in the pipeline [A]. In 2018 the global market was estimated at USD135,000,000,000, with further growth to over USD200,000,000,000 anticipated [B].

However, the generation of novel antibody therapeutics is currently still primarily a very time-consuming and expensive experimental endeavour. Pharmaceutical companies have therefore begun using databases and computational tools to accelerate the process of engineering an antibody to bind to a particular target or identifying antibodies that are markers for exposure or measures of vaccine efficacy. The SAbDab-SAbPred suite contains leading-edge tools that have provided this capability rapidly and efficiently. The resulting economic impact is realised by over 100 companies that have exploited this suite of tools.

Pathway to Impact

Deane took an early decision to make the SAbDab-SAbPred suite freely available under an open source licence in order to maximise the usage and impact of the tools. Funding from an EPSRC Impact Acceleration Award was used in 2016 to hire a Research Software Engineer (RSE) to further refine and integrate the software tools, making them easily accessible through a web server, and also to assist pharmaceutical companies who wished to bring the tools in-house and wanted help in adding new custom features. In 2020, to make it even easier for companies to use the tools, a second RSE was hired and the SAbDab-SAbPred suite was packaged as a virtual machine, named SAbBox, that can be downloaded as a single encapsulated licensed product.

9 pharmaceutical companies (GSK, UCB, Roche, AstraZeneca, Lonza, Iontas, Kymab, Tikcro, and Agenus Bio) currently have custom installations of the SAbDab-SAbPred suite. Five companies have already purchased SAbBox licences: GlaxoSmithKline plc (GSK), Kotai Biotech, Gritstone Oncology, Cenmed Enterprise, and Dragonfly Therapeutics. These collaborations have resulted in staff from AstraZeneca, Roche, GSK and UCB being involved as co-authors in much of the underpinning research [1,3,4,5,6].

Impact of SAbDab-SAbPred

The SAbDab-SAbPred suite has been utilised by over 100 pharmaceutical companies across the globe as part of their antibody therapeutic development pipelines [C]. These companies include very large multinationals such as Abbott, Medimmune/AstraZeneca, Bristol-Myers, Genentech, GSK, Novartis, Roche, Sanofi, and UCB as well as many smaller companies including Just Evotec Biologics and Voyager Therapeutics.

Components of SAbDab-SAbPred and/or the underlying research papers have been cited in 71 different patents (including 3 from Roche and 1 from UCB) in the area of therapeutic development [D], with some patents being filed in multiple regions (e.g. Europe, USA, Japan, China). Source [D] also includes the full text of 4 of these patents; two by Roche and UCB cite the use of SAbDab, while the other two by Seattle Genetics and Humabs BioMed cite ANARCI. The impact of using SAbDab-SAbPred by pharmaceutical companies has been to enable faster and more accurate results than commercial tools, thereby significantly reducing development costs and avoiding unnecessary expenses. The benefits that have been realised are illustrated below by company representatives of two large multinationals, GSK and UCB, along with two SMEs, Just Evotec Biologics and Voyager Therapeutics. Each outlines the particular impact that the suite of tools has had on their activities.

GlaxoSmithKline plc (GSK) is a British multinational pharmaceutical company with around 100,000 employees. It has collaborated with OPIG from its inception in 2008. A GSK employee was a co-author on publication [5], and GSK has paid for a custom installation of SAbDab-SAbPred and also a licence for SAbBox.

The use by GSK of these University of Oxford tools has been of great economic benefit. A GSK Scientific Leader & GSK Fellow in Protein Design and Informatics, Data & Computational Sciences [E] states: "*We use ABodyBuilder as our main antibody homology modelling tool, which in-house benchmarking has shown to be an order of magnitude faster, and of higher accuracy, when compared to the antibody modelling tool in* [the software package] *MOE. High quality homology models are required for many subsequent activities, such as the investigation of experimentally determined developability risks and identification of residues for mutation. ABodyBuilder is also embedded in our in silico assessment pipeline, as homology models are a pre-requisite for our in-house post-translational modification and isomerisation prediction methods… the prediction and mitigation of developability risks during discovery reduces the cost of development by an estimated £500,000* [GBP500,000] *to £2,000,000* [GBP2,000,000] *in wet work* [experiments] *and FTE* [staff] *per campaign depending upon the number of liabilities."* GSK indicate [F]*: "unfortunately the number of molecules* [campaigns] *we put through to development is not something we would disclose externally."*

UCB is a multinational bio-pharmaceutical company headquartered in Brussels, with around 7,500 employees. Like GSK, it has been involved in OPIG from its inception with a paid custom-installation of SAbDab-SAbPred. UCB has co-authors on publications [1,3,4,5,6].

At UCB, SAbDab and SAbPred have become an integral part of their antibody therapeutic discovery pipelines and have provided them with significant savings. The Senior Director, Head of Global Computer-Aided Drug Design at UCB [G], says of SAbDab*:* "*This comprehensive antibody structure database provides the data foundation for UCB's computational antibody design. The antibody design team within my department would not function without this data source. Should my team lose access to SAbDab, we will have to recruit a full-time contractor for data curation and database maintenance. The expected cost would be £150,000* [GBP150,000] *per year."* Concerning SAbPred, they say [G]: *"The toolbox is used routinely by the antibody design team in my department to support every antibody therapeutic project at UCB. This toolbox is indispensable to our antibody optimization workflow and, in our hands, produces more informative results than commercial toolkits offered by two well-recognized software companies specialized in computational drug discovery. Each of the commercial toolkits would cost over £250,000* [GBP250,000] *annually for the size of UCB."*

With their database, tools and an understanding of company needs, OPIG were ideally placed to respond when UCB commissioned a study to rapidly evaluate software developed by Massachusetts Institute of Technology. The Senior Director states [G]: "*In July 2020, UCB wanted to evaluate the feasibility of adopting an open-source AI software developed by Massachusetts Institute of Technology and published in Science. Within a month UCB received from your team not only an annotated, enhanced version of the open-source software, but also a report detailing components/functions that are missing from the software. This analysis led to UCB's conclusion that it will take £450,000* [GBP450,000] *and 1 year to adopt this AI software. Therefore, UCB invested in an alternative solution which costs £130,000* [GBP130,000] *and 2 months to implement. The savings of £320,000* [GBP320,000] *and 10 months, thanks to the Impact Software Engineer Service which costed UCB only £40,000* [GBP40,000]*…".*

Just Evotec Biologics is a Seattle-based biotech company focussed on bio-therapeutics, with approximately 200 employees. Just Evotec Biologics rely on tools from the SAbDab-SAbPred suite, using them to assess developability of antibodies on important development campaigns including an anti-HIV antibody. Use of these tools has led to the company realising huge savings. A Principal Scientist at Just Evotec Biologics [H] states: *"…we rely on tools such as those provided by the Oxford Protein Informatics Group (OPIG)… Optimization with respect to*

developability has been shown to have a profound beneficial impact on conformational stability… Additionally, well behaved, developable antibodies can be used… saving tens of thousands of dollars [USD10,000] *and months of process development time during the development process alone and ultimately leading to lower drug costs for patients. Most recently, this work benefited the optimization and process development of an anti-HIV bnAb with a potential global impact".*

Voyager Therapeutics is a Boston-based biotech company with around 200 employees. They have replaced commercial tools with ABodyBuilder, TAP and ANARCI to realise significant financial savings. A Senior Scientist in Vector Engineering, Antibody Discovery / Design at Voyager Therapeutics [I] states they have used ABodyBuilder, TAP and ANARCI: *"…to predict antibody homology models which saves substantial cost (>$100K/Ab* [more than USD100,000 per antibody]*) towards experimental structure determination*…[and have]…*replaced huge licenses costs incurred from Schrodinger, Chemical Computing Group and others ($50K/yr* [USD50,000 per year]*)."*

In summary, research from OPIG has enabled and accelerated antibody therapeutic discovery in pharmaceutical companies, both large and small, saving both time and money, reducing the costly experimental effort required, and bringing treatments to patients faster.

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

[A] Antibody Society webpage and spreadsheet of approved antibodies [accessed 31 Dec 2020]: https://www.antibodysociety.org/resources/approved-antibodies/

[B] BusinessWire website article on Monoclonal Antibodies Market Report 2020 [accessed 31 Dec 2020], https://www.businesswire.com/news/home/20191211005627/en/Global-Monoclonal-Antibodies-mAbs-Market-Report-2020

[C] List from OPIG, Department of Statistics, University of Oxford, of industrial users of SAbDab, SAbPred and OAS, 2020.

[D] List of patents citing OPIG tools and papers, and full details of 4 patents citing SAbDab or ANARCI, 2020:
i. Hoffmann-La Roche, WO-2016062734-A1: Vh-vl-interdomain angle based antibody humanization; filed 21/10/2015; https://patents.google.com/patent/WO2016062734A1/en
ii. UCB Biopharma, WO-2017093435-A1: De novo antibody design; filed 1/12/2016; https://patents.google.com/patent/WO2017093435A1/en
iii. Seattle Genetics, WO-2020117373-A1: Pharmaceutical compositions comprising anti-191p4d12 antibody drug conjugates and methods of use thereof; filed 15/10/2019 https://patents.google.com/patent/WO2020117373A1/en
iv. Humabs Biomed, WO-2020132091-A2: Antibodies that neutralize hepatitis b virus and uses thereof; filed 18/12/2019 https://patents.google.com/patent/WO2020132091A2/en

[E] Letter from GSK - Scientific Leader & GSK Fellow, Protein Design and Informatics, Data & Computational Sciences, January 2021.

[F] Email from GSK - Investigator, Computational Antibody Engineering, Protein Design and Informatics, January 2021.

[G] Letter from UCB - Senior Director, Head of Global Computer-Aided Drug Design, October 2020.

[H] Letter from Just Evotec Biologics - Principal Scientist, Molecular Design, April 2020.

[I] Letter from Voyager Therapeutics - Senior Scientist II, Vector Engineering, Antibody Discovery/Design, March 2020.