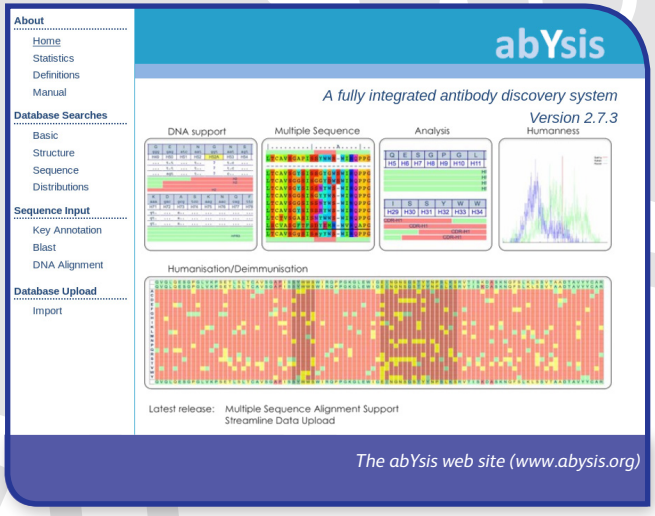
# Pharmaceuticals industry benefits from antibody sequence software

BBSRC funding has contributed to the development of antibody sequence analysis software abYsis, developed by Professor Andrew Martin at University College London. The software is currently licenced to 14 pharmaceutical industry users, and the free on-line version is regularly accessed by academic researchers and more than 150 industry research groups.

abYsis allows researchers to compare their antibody sequences with those in publicly-available libraries . It is used in antibody engineering, and is particularly useful for pharmaceutical and biotechnology companies seeking to design new antibody-based medicines.

The software arose from Martin's research on antibody sequence and structure and his efforts to develop software to benefit others working in the area.

"I've always been interested in doing research that is useful to somebody and that has a clear application," says Martin. "Developing software tools to help others has always been one of my main aims."




*The abYsis web site (www.abysis.org)*

## Comparing antibody sequences

The original version of abYsis came from an industry-funded research project supported by Celltech (now part of UCB). Martin was working with the company as an expert witness in a series of related patent disputes. At that time, "there was no sensible resource for storing and comparing antibody sequences," says Martin.

To create abYsis, Martin needed a method to label and compare antibody sequences consistently. Antibodies are proteins, built from a specific sequence of amino acids that fold in a particular way. The resulting three-dimensional structure gives antibodies their prized ability to bind to specific target molecules. Several schemes to number the amino acid sequences that form antibodies already existed, and a BBSRC-funded PhD student at UCL had developed an automated approach to apply those numbering schemes, which Martin could incorporate into abYsis .

Subsequent follow-on funding from BBSRC enabled Martin to update abYsis, based on feedback from users. The funding supported development of an improved online user interface, as well as enhancing abYsis to work with unusual antibody sequences from species such as sharks or camels.

## Unique or unusual sequences

"About one third of all drugs in development are based on antibodies," says Martin. In nature, antibodies play a key role in our adaptive immune system. Their specific shapes allow them to bind to a vast range of targets, either neutralising the target directly or tagging it for destruction by other elements of the immune system. This specificity, combined with researchers' ability to modify antibody structure to alter or improve their function, means that antibodies can also be harnessed in medicines to target and treat diseases.

abYsis allows users to compare their antibody sequence data with sequences stored in databases such as the Protein Data Bank , IMGT , or GenBank , together with structural data from the Protein Data Bank.

Comparing sequences allows researchers to identify unique or unusual features of their antibody sequences. It also

## IMPACT SUMMARY

**Dr Andrew Martin and colleagues at UCL developed antibody sequence software abYsis, which is being used by more than 150 industry groups to analyse the sequences of amino acids that form antibodies, often as part of their drug development programmes. Antibodies can be tailored to identify and bind to specific targets, and pharmaceutical companies are interested in using them as the basis of new medicines.**

**The software arose from an industry-funded research project and has subsequently benefitted from BBSRC investment, including a BBSRC-funded studentship and follow-on funding.**

**Industry users range from small biotech firms to large multinational pharmaceutical companies. abYsis allows users to compare sequence data with that from public libraries of antibody sequences via a free-to-use web interface. Users can also buy software licences to use abYsis in-house. There are currently 14 licence-holders, and the income from these means abYsis is now largely self-funding.**

## SUPPORTING THE PHARMACEUTICALS INDUSTRY

**As part of the evidence for a 2014 REF case study, Martin gathered supporting statements from industry users of abYsis, including:**

**Pfizer: "the ability to use [the software] to search large, well annotated databases of antibody sequences and to accurately annotate or otherwise analyse new proprietary sequences that Pfizer develops over time, is important to Pfizer"**

**UCB: "Abysis is a tool that brings external information into a more digestable form, this has the following effects for us internally; generally it is about better informed processes, through delivering public information into the value chain, the main area of support is in the decision making around antibody engineering".**

allows them to determine how human-like the antibodies appear. This is important when designing antibody-based therapeutics, as our immune system will target unfamiliar molecules, rendering medicines based solely on non-human antibody sequences ineffective and potentially leading to severe side-effects such as anaphylactic shock. Pharmaceutical companies therefore need to humanise their antibodies, which are often produced in mice. "The abYsis software helps them do that," says Martin.

### Industry licences and online users

To commercialise the software, Martin enlisted the help of Dr Mark Swindells.

"I knew what abYsis did, I knew there was not much antibody software available at the time, and I knew lots of people were accessing the public version," says Swindells.

Following a conversation with UCL Business, who supported the idea, Swindells began to take simple steps to raise the profile of abYsis and to target potential commercial users. "Academic research doesn't have to deliver commercial opportunities," says Swindells. "But when the opportunity is there, it's a shame when it isn't taken further."

By talking to existing contacts, Swindells established three commercial licences with companies in the first twelve months. Since then the growth has been carefully managed to allow Martin and Swindells to learn from each new user and make further improvements to the software.

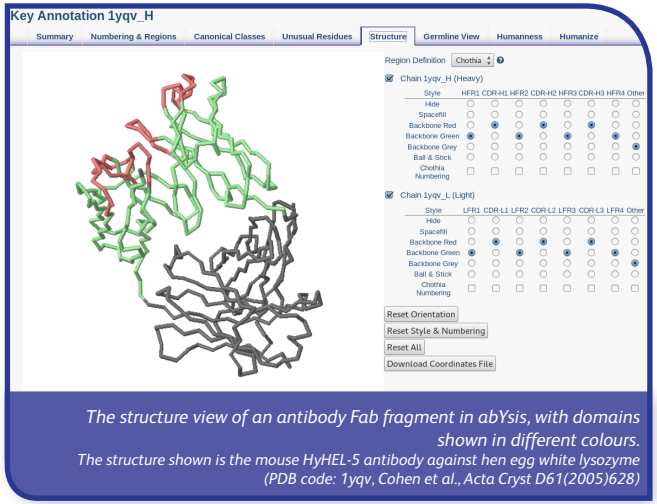
abYsis is now largely self-funding, and employs two developers to work on the software full-time. By 2018 fourteen companies held commercial licences for abYsis, to use the software in their antibody therapeutics development programmes. Companies pay for the licence as it allows them to use the software in-house to analyse their own sequences without transferring their highly confidential data onto a public server.

Typically, a licence costs between £10,000 and £25,000 per year, depending on the number of people who need access. Licencees range from small biotech firms to major multinational companies, and come from all over the world. Due to their initial support for the project, UCL have granted UCB a licence in perpetuity.

The software is also free to use online, and is being widely used by academic researchers and companies; over 150 commercial groups have now used the software for free.

### Next steps: One million sequences

The next step for abYsis is to increase the number of sequences it can handle at once, as pharmaceutical companies are increasingly interested in analysing very

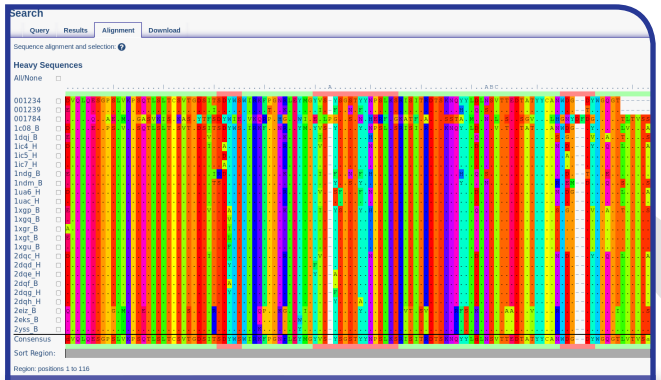



*The structure view of an antibody Fab fragment in abYsis, with domains shown in different colours.*
*The structure shown is the mouse HyHEL-5 antibody against hen egg white lysozyme (PDB code: 1yqv, Cohen et al., Acta Cryst D61(2005)628)*

large numbers of sequences held in antibody libraries. Martin is now looking to build a partner system using a different type of underpinning database software that is specifically designed to handle millions of data points. "We went to visit a company in Cambridge," Martin explains. "They asked if the system could deal with one million sequences which we said would be no problem, but they went on to say they were generating one million sequences per week. The system as it was at the time, took about two weeks to get one million sequences into the system, and that might be a bit impractical!"

"For dealing with that sort of thing we want to move to one of the new no-SQL databases," he adds. "These are designed to deal very quickly with huge datasets and the idea is to build a companion system that can be used to store and filter these enormous datasets before feeding the sequences of interest into abYsis."

"The system can process larger datasets," says Swindells, "but how does a user process one million datasets? What strategy do they use to get the number down?"




*A sequence alignment of antibody heavy chains shown in the JSAV alignment viewer within abYsis*
*Image: JSAV is available as open source code from http://www.bioinf.org.uk/software/jsav/ (Martin, A.C.R. F1000Research 3(2014)249)*

To address user needs, abYsis needs to be able to help them answer the specific questions they want to ask, while being sufficiently generic that it can be used by multiple users. To help users, abYsis now has a more attractive front end that enables quicker and easier access to the data, and the abYsis team are continuously learning from their users to help adapt abYsis to meet their evolving needs.

"Early adopters wouldn't recognise the system now, compared to the early system when they took the first licences," Swindells adds.

## REFERENCES

1 Dr Andrew Martin, UCL: [Reference/webpage no longer available – September 2018]
2 abYsis: http://www.abYsis.org/
3 Swindells MB, Porter CT, Couch M, Hurst J, Abhinandan KR, Nielsen JH, Macindoe G, Hetherington J, Martin ACR. (2017). abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction. J Mol Biol. 429(3):356-364. doi: 10.1016/j.jmb.2016.08.019.
4 UCB: https://www.ucb.com/