

Andy Lau, Moleculomics



'My programming skills have definitely improved. Resuming my own PhD project now, I feel like there are certain problems that I can solve easily with coding compared to before.'

Whilst at Moleculomics I was involved in developing a computational methodology that was able to benefit the company. Moleculomics's flagship product is the Human3DProteome platform, which contains the results of millions of simulated drug screening trials across every protein in the human genome.

This information is accessible in a large comprehensive database, backed up by quantitative scores and data from other non-computational sources. As the company deals with computational drug screening, it is imperative that the protein models used for docking drug compounds, are accurately represented (this is a major topic in academia as well as industry). My project was focused on developing neural network algorithms that are able to predict how proteins might be able to complex together.

I was employed by Moleculomics as a Researcher in Structural Bioinformatics, which accurately described my roles within the team. Working alongside another developer, I designed scripts and programs used to gather data points in a large dataset and then analyzed these using different machine-learning methods.

Overall, I developed strong practical skills in not only Python, but also on how to implement machine-learning methods such as neural networks and support vector machines. The project will hopefully lead to an interesting academic output in which for the first time, protein complexes can be computational generated to high accuracy.

Because of the computational nature of the project, and my lack of computational background, I expected that there would be a bit of a steep learning curve at the beginning of the project. However, in reality, I approached the project first using MATLAB, which I learnt through LIDo, before, and then converted my scripts into Python. Since then, I have gained a solid foundation in Python and I think overall, the project has allowed me to develop some extremely applicable and valuable computational skills.

While I think this project could have been tackled by someone without my biochemistry/structural biology qualifications, I think my prior knowledge in these areas were

useful in several key points in my projects. There were times when I needed to decide how to gather data for the machine learning programs, and I think I was only able to solve these issues with my developed intuition of structural biology.

While I have used Python before, I always defaulted back to MATLAB due to it being my most familiar programming language. During and after the placement, Python has taken over as my main language and I am very grateful for this as Python is one of the most useful and adaptable languages. Part of the reason why I decided on computational PIPS was also because I wanted to force myself to have to learn Python.

I will certainly stay in contact with Moleculomics and have decided to continue with the project until it has been finalized – post placement. We will continue to have periodic Skype meetings to discuss progress.

My programming skills have definitely improved. Resuming my own PhD project now, I feel like there are certain problems that I can solve easily with coding compared to before. It was quite interesting to see how research happens in a commercial setting - the biggest difference I noticed was how developers pick up projects, finish them, and move on, rather than continuing with ongoing long term projects. Compared to academic research, this seems much more efficient and develops your project management skills because you are exposed to many more projects while needing to work more efficiently. I think I am better suited for data scientist roles and this is an option that I would be interested to look at in the future.

To find our more contact Andy Lau directy: andy.lau.15@ucl.ac.uk