Biomolecular Simulations and Calculations | Dr Andrew Torda

In 1995, the group entered a phase of exponential growth, moving from one to two people. In 2002, the group shrunk equally swiftly, finishing with zero members by the end of September.

During the last nine months of its existence the small, but merry band (M.! Abraham, Zs. Dosztányi, J. Procter) continued to work on trying to improve methods for guessing a protein's structure given only its sequence. As in the past, this was done by creating and parameterising functions which act like energy terms, but have lost their connection to real physics. This is not done simply to be obscure. Instead, there is an underlying philosophy that one should optimise performance on a smaller problem at the cost of generality. In our case, this means working on methods which are very adept at recognising good pieces of proteins and trying to tell correct from incorrect structures. The cost is that they cannot tell wrong from very wrong. In the right context, this is a powerful computational tool.

For the technically minded, there was a wholesale rethink of our approach. For some years we were stuck with our baggage from molecular mechanics backgrounds and parameterised functions which were largely based on pairwise, through-space interactions. In the last year, we were seduced by ideas from Bayesian statistics which is somewhere between a statistical formalism and religious dogma. In practice, we built (along with our University of Queensland collaborators) functions which treated protein sequence and some structural properties as equal statistical descriptors and resulted in a very general measure of sequence to structure compatibility. The nomenclature is different from much of computational chemistry, but is really simply making something explicit which is always underlying a chemist's thinking – how appropriate is some structure for some set of atoms or particles or amino acids.

We also undertook a complete rewrite of our older structure prediction code. This was travelling under the silly name of sausage. It now has the more sensible name of "wurst". To the astonishment of all concerned, this work was finished on schedule and in time for an international comparison of protein structure prediction methods held at the end of 2002. We did achieve our main aim of having the code so modular, we could have one program which could be used for parameterisation, testing and application to unknown cases.

Although building force fields or score functions is fun, it is also interesting to see if they can be put to other uses, and they certainly can. We have always worked on functions which are based on extracting information from known protein structures. Obviously these can be biased by the selection of proteins used for parameterisation. For example, if you use a set of large proteins, you will get different results from a score function built from small proteins. We have tried to turn this phenomenon to our advantage. In 2001 we developed a method for extracting the dominant features from a low-resolution force field. This year, we applied this to score functions deliberately constructed from special sets of proteins. This means one gets quantitative estimates of the factors which are important for small proteins or transmembrane proteins, for example. Continuing in this vein, we have begun to use score functions to characterise and compare proteins. This is an interesting problem since known protein structures are not easy to compare to each other. Proteins are three dimensional objects of different

size where there is no obvious recipe for comparison, despite the ability of humans to say this one looks like that one. To use the terminology from physics, one can say that proteins provide a field and, with our score functions this can be swiftly calculated. The behaviour of some kind of particle within this field provides a fingerprint of a protein and since proteins can be seen as linear polymers, this can be treated as a string. With this kind of approach, one is able to use classic string comparison algorithms to recognise similarities. Reassuringly, this provides results similar to what one would see by looking at protein structures. Most importantly, we have used this as the basis for clustering of proteins structures.

The group lived in the Research School of Chemistry for just over seven years. Although reports like this one mention its activities alone, the group drew heavily on its environment for spiritual guidance and it borrowed heavily in areas such as numerical methods. It was known as the Biomolecular Simulations group in Canberra and bears a striking resemblance to something known as the Gruppe für Biomolekulare Simulations in a north German university.

http://www.zbh.uni-hamburg.de/mitarbeiter/torda.htm