Janet Thornton

Bioinformatician *avant la lettre*

Michael Gross

Janet Thornton

Bioinformatics is very much a buzzword of our time, with new courses and institutes dedicated to it sprouting up almost everywhere. Most significantly, the flood of genome data has raised the general awareness of the need to deve-lop new computational approaches to make sense of all the raw information collected.

Professor Janet Thornton, the current director of the European Bioinformatics Institute (EBI), an EMBL outpost based at the Hinxton campus near Cambridge, has been in the field even before there was a word for it. Coming to structural biology with a physics degree from the University of Nottingham, she was already involved with computer-generated structural images in the 1970s, when personal computers and user-friendly programs had yet to be invented.

The Early Years

From there to the EBI, her remarkable career appears to be organised in decades. During the 1970s, she did doctoral and post-doctoral research at the Molecular Biophysics Laboratory in Oxford and at the National Institute for Medical Research in Mill Hill, near London. The former, then under the leadership of David Phillips is famous for its protein crystallography work, the latter for structural NMR. Thus Thornton was involved with the elucidation of protein structures at a time when the explosive growth of structural biology was only beginning. During the 1980s, she worked at the department of crystallography at Birkbeck College, London. Although she also started raising a family during this time and was employed part-time for a number of years, she became a lecturer and eventually a senior lecturer at Birkbeck, and established her reputation in structural biology by analysing common motifs in the growing number of known protein structures. During that time, it began to emerge from her work that, while the number of structures increased exponentially over time, the number of truly different arrangements of the polypeptide chain, the "folds", remained relatively small and might be limited to a surprisingly small number, perhaps less than 1000. Her group started developing software to compare structures to each other, recognise known folds and spot new ones. Such work provides both fundamental insights into the workings of evolution on a molecular level, and practical benefits, as it facilitates the prediction of structural and functional aspects from sequence data.

University College

In 1990, Thornton moved to University College London (UCL), just a few hundred metres down the road from Birkbeck. UCL offered her a chair dedicated exclusively to what we now call bioinformatics. The new position enabled her to develop her specific computational approach to structural biology to full bloom. Programs developed by her lab are used by researchers world wide. One particularly useful piece of software, now incorporated in platform known as BioPendium, led her to the foundation of a spinout company, Inpharmatica, in 1998.

Biopendium is a tool to annotate new sequences. It picks up all the publicly available sequence data and all of the more than 30,000 structures, and allows the user to compare all of them, and to find even distant relationships or similarities. Within 15 minutes, the software can check all 2.4 billion possible relationships and pick the ones relevant to the question at hand. In comparison to publicly available bioinformatics packages such as Blast or Psiblast, Biopendium can provide an additional 30 % of annotation, according to Inpharmatica's managers. For example, public domain searches for kinases would give you less than 800 examples, while Biopendium came up with 973.

Launched in 1999, Biopendium was successful enough to allow Inpharmatica to grow and pursue an unusual two-thronged business strategy. The company continues to sell data resources like Biopendium, but it is also involved in drug discovery collaboration based on their second big development, Pharmacarta, which matches proteins to small molecules, and which is available only on the basis of collaborative agreements. Inpharmatica has forged such partnerships with companies including Swiss-based biotech company Serono. While the company has successfully detached itself from its academic origin in the UCL research labs – it boasts IT-heavy corporate headquarters in Charlotte Street, West London, along with research sites in Camden and Hemel Hampstead – Thornton keeps a close eye on it, as she chairs the scientific advisory committee.

BIOforum Europe 4/2003, pp 178–179, GIT VERLAG GmbH & Co. KG, Darmstadt, Germany, www.gitverlag.com/go/bioint

Another offshoot of Thornton's work at UCL is the CATH database of structural domains, a systematic classification of folds based on the hierarchical criteria of Class, Architecture, Topology, and Homologous superfamily. This project has remained at UCL, where a team of around 20 PhD students and post-docs, supervised by Christine Orengo, keeps it going **(http://www.biochem.ucl.ac.uk/bsm/cath_new/)**.

During the last years at UCL, Thornton also chaired two initiatives shared with her former department at Birkbeck, namely the Joint Research School in Biomolecular Sciences and the BBSRC-funded Bloomsbury Centre for Structural Biology.

EBI

Since 1994, she has also been involved with the EBI as a consultant, and finally became the institute's director in October 2001, with initially a five-year tenure. While she keeps up her involvement with fundamental research, the new job is very much that of a mover and shaker. The exponentially growing flood of of biological data requires strategic planning. As Thornton puts it: "we need a 5-year plan that will involve some expansion, including the need for additional space." Part of her brief is to expand the activities of the EBI, which currently maintains a number of public-domain databases with gene and protein sequences, structures, and expression data. Using this recognised strength of the institute as a foundation, she is building a stronger research institute, with teams from different disciplines interacting with each other and with the expertise of the database facilities. New group leaders have already been appointed, and further positions will be advertised in the autumn. For this expansion, the core funding that the EBI receives from the EMBL has been increased to € 10 million. On top of that, there are major grants from the Wellcome Trust and the European Union, among others. The growing enterprise of bioinformatics also needs connections to the outside world. Thornton says: "We will need to provide bridges between the biomolecular data we collect and archive, and the related data in chemistry, medicine or other important areas of application."

On joining the EBI, Thornton has also emphasized the need to develop interfaces with smaller biotech companies, as well as the importance of spreading expertise in bioinformatics. "Training is a very important part of the role of the EBI and we need to complement the provisions made by the universities in Masters and PhD courses. Therefore our focus is very much to train the trainers throughout Europe. We are hoping to establish a forum for the scientists who teach bioinformatics to share experiences and to provide updating on new technologies in bioinformatics. This forum is still in the planning stage."

Meanwhile, Thornton chairs the organising committee for a major bioinformatics conference to be held in Glasgow, August 2004. The third European Conference on Computational Biology will be the first in the series to be held in the UK, and the first to be held in conjunction with the leading annual international bioinformaices conference series on Intelligent Systems in Molecular Biology. It looks like bioinformatics will keep growing bigger for some time to come.

Weblinks:

www.cryst.bbk.ac.uk Birkbeck Crystallography
www.biochem.ucl.ac.uk UCL Biochemistry
www.inpharmatica.co.uk Inpharmatica
www.ebi.ac.uk EBI

Michael Gross
He is a science writer in residence at the school of crystallography, Birkbeck College, University of London. His latest book, "Light and Life" is available from Oxford University Press.
See www.proseandpassion.com for details.