

## Summary of Completed Project

The specific aims of the Third International Conference on Intelligent Systems for Molecular Biology (ISMB-95) were to:

- convene a critical mass of researchers applying advanced computational techniques to problems in molecular biology
- promote interchange of problems and solutions between computer scientists and molecular biologists
- create education opportunities in this cross-disciplinary field for students and senior researchers wishing to either apply or benefit from these techniques
- produce an archival proceedings as a forum for rapid dissemination of new results in a peer-reviewed manner.
- produce a set of tutorial materials for education and training of researchers interested in this field
- maintain the momentum generated by the highly successful previous conferences in the series, and establish a regular event that will help to solidify the field
- foster the involvement of women and minorities in the field

## Technical Information

Third International Conference on Intelligent Systems for Molecular Biology,  
Robinson College, Cambridge, UK. July 16-19, 1995.

The principal aim of ISMB-95 was to bring together scientists who are addressing problems in molecular biology using advanced computational methods including data modelling, machine learning, artificial intelligence, cognitive science, robotics, combinatorial/stochastic optimization, adaptive computing, string & graph algorithms, linguistic methods and parallel computer technologies. The ISMB conferences have been probably the most prominent and successful gatherings for this community and the only conference with a commitment to being international. ISMB-95 followed two successful conferences held in 1993 and 1994 in Bethesda and Stanford, USA. In terms of numbers of delegates, however, ISMB-95 was twice as large as ISMB-93 and 1.5 times larger than ISMB-94. The conference, which was fully subscribed, attracted 269 delegates, about 150 of which were from the US, 60 of whom were from the UK, with around 100 from other countries.

ISMB-95 received support to fund the attendance of young scientists as a result of contributions from the National Science Foundation, Department of Energy, and National Institutes of Health in the US (34 delegates). It also received funding from the Medical Research Council (3 delegates), Biotechnology and Biological Sciences Research Council (10 delegates) and CEC Euroconferences programme in the UK. Finally, support was provided by the Imperial Cancer Research Fund, and we were in

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

receipt of commercial sponsorship from Glaxo Wellcome PLC and Oxford Molecular Group PLC.

The first day of the conference was taken up with a programme of 8 introductory and advanced tutorials, attended by 187 delegates in total, and given by respected scientists:

Dr Shankar Subramaniam, Beckman Institute, USA

(Molecular biology for the computer scientist)

Professor Doug Brutlag, Stanford University School of Medicine, USA

(Intelligent systems for the molecular biologist)

Dr Sean OeDonoghue/Dr Burkhard Rost, EMBL, Heidelberg, Germany

(Protein Structure Prediction)

Dr Pierre Baldi, CALTECH, USA

(Machine learning approaches to molecular biology)

Professor David Searls, University of Pennsylvania School of Medicine, USA

(Linguistics of biosequences)

Dr William Pearson, Dept. of Biochemistry, University of Virginia, USA

(Protein Evolution)

Dr Thomas Lengauer, GMD-II, Germany/Dr Gerhard Klebe, BASF, Germany

(Computational methods for molecular docking)

Dr Chip Lawrence, Wadsworth Center for Laboratories and Research, USA

(Statistical foundations of multiple sequence alignment and structure prediction)

The level of attendance represented a marked increase on the numbers attending tutorials at ISMB-94 and is indicative of the important complementary role of the tutorial programme to the conference as a whole. The numbers attending the 2 introductory tutorials (aimed at giving computer scientists an introductory knowledge of molecular biology, and at introducing molecular biologists to relevant computational techniques) were relatively lower than at the previous 2 conferences - perhaps an indication that computational molecular biology is now fully established as a distinct interdisciplinary research and development area - but there was still a satisfactory level of interest from researchers keen to extend their experience across the disciplines. The tutorials which attracted the highest number of delegates (approx. 70 each) were on Protein Structure Prediction and Statistical Foundations of Multiple Sequence Alignments and Structure Prediction.

The popularity of the tutorial programme shows that there is a large demand for high quality training in this area. The level of interest in the tutorial programme was such that we sold out of most of the individual tutorial booklets available, and are still receiving requests to purchase these from delegates and non-delegates alike. All the delegates who participated in the tutorial programme and we have encouraged them to provide us with short written reports on their impressions of the effectiveness of the tutorial programme.

In addition to the peer-reviewed presentations, plenary lectures were given by distinguished scientists Walter Bodmer (Imperial Cancer Research Fund, UK), Douglas Brutlag (Stanford University, USA), John Sulston (Sanger Center, Cambridge, UK), and Janet Thornton (University College, London, UK).

The conference proper consisted of 8 sessions covering key bioinformatics issues: Protein Structure and Docking, Sequence Alignment, Protein Sequence and

### Structure,

Understanding Sequence Function, RNA Sequence and Structure, Genome Information Systems, Gene Finding and Gene Structure and Database Searching (26 papers in all). Each session comprised 3 or 4 oral presentations of papers which were included in the conference proceedings only after a rigorous peer review process (undertaken by members of the conferencees Programme Committee), and each session was chaired by a member of the Organizing Committee with a special knowledge of the topic under discussion. Once again, we have asked various recipients of grant funding to provide us with reports on individual sessions.

Issues which seemed to raise particular interest this year included methods for automatically identifying the different structural and functional domains within protein structures and the use of advanced statistical methods (e.g. hidden markov models and stochastic context-free grammars) for identifying complex protein sequence patterns and sequence relationships. The integration of genome information and its presentation to users was a theme which recurred in the oral and poster sessions, as were methods for rapidly searching databases of DNA sequence and protein sequence and structure for related molecules.

A new topic which appeared this year was the search for efficient and accurate methods for finding the best fit between macromolecular structures (docking) - both protein-protein and protein-ligand interactions being considered. Several papers (and a tutorial) covered the problems of protein structure prediction and refinements to methods for threading sequences through known protein structures. The application of machine learning methods (e.g. neural networks, hidden markov models and genetic algorithms) were well represented and a new approach based on a simulation of the immune system was presented. RNA structure prediction was addressed by several different techniques including the use of linguistics techniques and graph theory.

Throughout the meeting, however, two general trends could be discerned: the need to evaluate new approaches against existing methods with the associated complexities of selecting the appropriate datasets and secondly the desire to make predictions of biological function which could be properly validated. This raises the general problem, also addressed in poster and oral paper presentations, of how biological function is represented in molecular biology databases and the clear need for more complete databases of biological processes and pathways which can be linked to the sequence and structure datasets.

The oral presentations were complemented by 2 poster sessions: the first, "official" session comprised 22 posters which had, like the oral papers, been accepted for inclusion in the published proceedings of the conference as a result of the peer review process. The second "open" poster session comprised 43 posters and was held with a view to giving as many researchers as possible the chance to talk about and publicise their work. This was of particular value to young researchers. Both sessions were extremely well attended and seemed to generate a lot of interest and debate.

From an administrative point of view the conference ran very smoothly. Robinson College was regarded as a good choice of venue and feedback from delegates has been extremely positive. A general opinion has been

expressed that the level of interest in, and perceived success of, ISMB-95 will be a significant contributing factor to the establishment of computational molecular biology as a recognised scientific discipline and that successive ISMB conferences will be able to build on this achievement.

St. Louis was selected as the site for ISMB-96, and it is clear that the conference now has enough momentum to be self-sustaining. Future proposals to federal agencies for grant support are likely to request mostly (or exclusively) funds for travel of young scientists

## Dissemination

The web page for ISMB-95 is <http://www.biu.icnet.uk/ismb/ISMB95.html>.

The proceedings were allotted 9 pages in the proceedings, which was published by the AAAI Press in a high quality, soft cover format matching the proceedings from the previous year (Available from AAAI Press, Menlo Park, CA ISBN-0-929280-83-0). The official reference is:

Proceedings of Third International Conference on Intelligent Systems for Molecular Biology (Rawlings, Clark, Altman, Hunter, Lathrop & Wodak, eds), AAAI Press, Menlo Park, CA.

The National Library of Medicine has recently decided to include the papers within the ISMB proceedings in the MEDLINE database, which promises to increase the stature and availability of these reports. A request is pending to the INSPEC database of engineering and computer science literature for a similar listing.

The tutorial notes are also available separately from the organizers, and represent useful collections of papers and notes for interested readers.

A copy of the proceedings and all tutorial notes are included with this report to the NIH, NSF and DOE.

## Summary of Travel Awards

The 1995 meeting received support for travel of US scientists from the National Science Foundation (\$13,000), the Department of Energy (\$13,000), and the National Institutes of Health (\$13,000, funding divided between the National Center for Human Genome Research and the National Library of Medicine). Awards were made to 35 individuals. Applications were sought specifically from young scientists, women scientists and minor scientists, although all applications were considered. Fifteen graduate students and undergraduates, all authors on accepted papers or posters were funded. Ten awards were made to post-doctoral fellows, and ten awards went to faculty members with demonstrated need, and with accepted papers or tutorial. We supported 9 women and 2 minority scientists. We noted a decline in applications for funding from minority scientists compared to 1994, perhaps because of the expense of travel to Britain. Only partial support was provided for most awardees, with a maximum grant size of \$1300 (awarded to 22 recipients). In addition, some of the grants were smaller, based on status as faculty or other senior scientists with demonstrated need, but not as dire as for students. These awards were usually approximately \$750, and were given to 13 recipients. Travel awards were made using the same criterion proposed for this grant, as reviewed in the methods section of this proposal.

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

#### **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**