

The logo for the Engineering and Physical Sciences Research Council (EPSRC). It features the acronym "EPSRC" in a bold, blue, sans-serif font. The letters are contained within a blue rectangular border that has a slight 3D effect, with a darker blue shadow on the right and bottom sides.

Engineering and Physical Sciences
Research Council

The Future of Bioinformatics: A Paradigm Shift in the Computer Sciences?

EPSRC Workshop

30 November – 1 December 2004

Radcliffe House, University of Warwick

The Future of Bioinformatics: A Paradigm Shift in the Computer Sciences?

Executive Summary

The aim of this workshop was to identify current problem areas in the field of bioinformatics and identify future research directions which might contribute to solving these problems. The conclusions from the workshop will be used, along with the views of the BBSRC and other stakeholders, to inform the EPSRC's business planning process in spring 2005, and to define future EPSRC funding strategy in bioinformatics.

The workshop included invited talks from the four EPSRC-funded Chairs in Bioinformatics and two other guest speakers. The breakout discussions identified three broad problem areas: lack of theoretical foundation in bioinformatics; making the most of data; and working across disciplines.

1. Introduction

The aim of this workshop was to obtain expert input into the development of the EPSRC's future funding strategy in bioinformatics. To this end, the four EPSRC-funded Chairs in Bioinformatics were brought together with a mixed group of peers from both academia and industry to discuss future perspectives in bioinformatics.

2. Presentations

The workshop opened on 30 November with talks from the four EPSRC Bioinformatics Chairs on the future of bioinformatics. Prof. Andy Brass (University of Manchester) discussed education issues in computing research, and the importance of Grid technologies and healthcare informatics. Prof. Jotun Hein (University of Oxford) concentrated on the challenges presented by the evolution of protein structure, gene structure and networks, and stressed the need to synthesise computer science, mathematics and statistics in bioinformatics. Prof. David Jones (UCL) described possible future scenarios in experimental techniques and data types, and offered a tentative timeline for the next twenty years in the field. Prof. Stephen Muggleton (Imperial College) noted a shift in bioinformatics from sequence analysis towards biological networks, and emphasised that systems biology must consider communities, not organisms in isolation.

Two further invited speakers carried on the same forward-looking theme on 1 December. Dr Philippe Sanseau (GlaxoSmithKline) stressed that GSK bioinformatics activity increasingly covers the whole spectrum of the drug discovery pipeline, and identified data integration as a top priority. Prof. Muffy Calder (University of Glasgow) talked on future modelling challenges for computer science in biology, and detailed a particular example concerning cell signalling pathways.

3. Workshop breakout sessions

Two workshop breakout sessions were held on the second day of the workshop. The first of these was used to define problem areas in bioinformatics, and the second looked at some of these problems in detail, with the aim of coming up with new approaches to tackle the barriers to progress in the field.

Three broad, cross-cutting themes emerged from the discussions in the two breakout sessions. The following tables show the general problem areas identified within each theme and also the suggestions made on how to overcome some of these barriers to progress. It is worth noting that some areas overlap, and that there is not necessarily one solution noted for each problem.

The outputs from the breakout sessions will be used, together with those from a recent BBSRC bioinformatics grant-holders' workshop, as the basis for discussion of future EPSRC bioinformatics funding strategy, in partnership with the BBSRC.

Results from a review of the joint BBSRC/EPSRC bioinformatics initiative, which ran from 1994 to 2000, will also be considered once the review is completed later this year.

A. Theoretical framework

Problem areas	Possible approaches/solutions
<ul style="list-style-type: none">Bioinformatics needs to grow beyond a service to a discipline	<ul style="list-style-type: none">Encourage work with a theoretical element and

<p>in its own right.</p> <ul style="list-style-type: none"> • Scientific questions in the area may not be as focussed and well articulated as they need to be. • Biologists' 'subjective' knowledge needs to be represented and formalised to inform algorithms and models. • There is often a lack of synthesis and predictive power in current modelling techniques: bioinformatics needs to move towards integrative modelling and systems biology, e.g., system identification. • Fundamental computer science research is also needed to establish the basis for applications. 	<p>theoreticians into the field.</p> <ul style="list-style-type: none"> • Focus on identifying key functional parameters in biological networks, as opposed to sequence data. • Validate models and algorithms with experimental data. • Evaluate different modelling approaches by setting a 'prediction challenge'. • Develop scalable algorithms and integrative models. • Concentrate on achievable, medium-sized challenges, not grand challenges!
---	--

B. Making the most of data

Problem areas	Possible approaches/solutions
<ul style="list-style-type: none"> • There is a vast array of data types of varying quality to be encompassed. Interoperability is often a challenge. • Missing data and uncertainty are key issues to be dealt with. • Provenance and meta-data are crucial. Data should be re-usable. • There is a need to move beyond data collection for its own sake. • Subjectivity is an issue in some areas, e.g., phenotype data. • Terminology differences across disciplines cause data problems. • Ownership issues are important – providers need to get due credit for their work. There may be ethical considerations in some areas, especially medical informatics. • Standard interfaces are 	<ul style="list-style-type: none"> • Identify which sorts of data are really needed, and tailor databases accordingly. • Integrate data collection and model testing • Make the most of synergies where they exist, for example, sharing ontologies and applying existing visualisation techniques to new areas. • Licensing arrangement to ensure that data owners get due credit in the re-use of their data. • Set up data standards group from community. • Develop flexible representations, which conform to standards but allow development.

currently lacking for end-users of bioinformatics tools.	
--	--

C. Working across disciplines

Problem areas	Possible approaches/solutions
<ul style="list-style-type: none"> • Communication across the discipline interface is still difficult for many people. • It can be difficult to define common objectives across the interface. • Career structure at the interface is unclear. • HEI departmental structure often doesn't help in building links across disciplines. • The RAE is perceived as promoting risk-adverse research. • Publication at the interface may not be valued as highly as work in 'core' disciplines. • Integration of computing with wet experiments is a challenge. For example, one half of the project may be mundane, and not engage expert interest. • Experimental design is sometimes neglected – statistical expertise is needed to ensure the significance of results. 	<ul style="list-style-type: none"> • Promote two-way collaboration as opposed to concentrating on enabling technologies in each discipline. • Promote work at the interface between systems biology and bioinformatics – develop interdisciplinary research centres. • Build new industrial collaborations. • Move out of RAE 'comfort zone'; encourage high-risk research. • Increase undergraduate training – BSc in Bioinformatics. • Develop 'honourable failure' as a positive in adventurous projects. • More discipline hoppers! • Promote communication between communities – appoint 'translators'. • Develop links to e-science. • Specifically encourage the development of medical informatics.

4. Final wrap-up discussion

After the workshop sessions, the four Bioinformatics Chairs were invited to lead an open discussion on the most exciting messages as well as challenges to come out of the workshop. Some of the main points made are summarised here.

- The range of models and ideas from computer science brought to bear in bioinformatics has been impressive, but needs to increase still further.
- There is a rapidly increasing challenge to computer science from quantitative biology.
- There is a tension between multidisciplinary and scientific focus.

- The role of the Bioinformatics Chairs in making things happen in the field is very important.
- It is extremely important to engage biologists in the computing side of bioinformatics, and for biologists to describe biological problems in a way accessible to computer scientists.
- There is a need for challenging questions to develop bioinformatics into a true, hybrid discipline.

Dr Elizabeth Hylton, the EPSRC Life Sciences Interface Programme Manager, brought the meeting to a close by describing some of the mechanisms the EPSRC currently has in place to support novel approaches to research, and by thanking delegates for their input into the EPSRC's future funding strategy in bioinformatics, to be developed in association with the BBSRC.

Dr Gavin Salisbury
Associate Programme Manager
EPSRC Life Sciences Interface Programme
February 2005