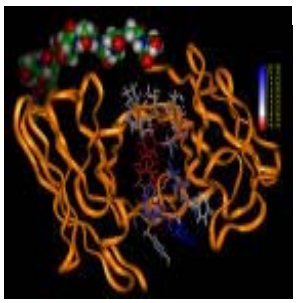
Integrative Biology

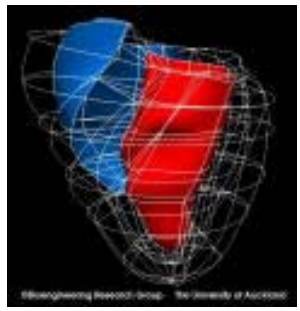
- exploiting e-science to combat fatal diseases



Tackling two Grand Challenge research questions:

- **What causes heart disease?**
- **How does a cancer form and grow?**

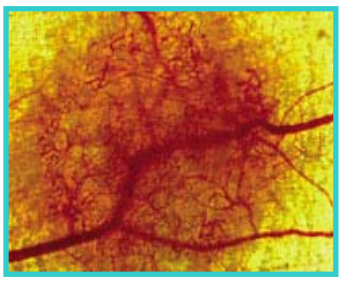
Together these diseases cause 61% of all UK deaths

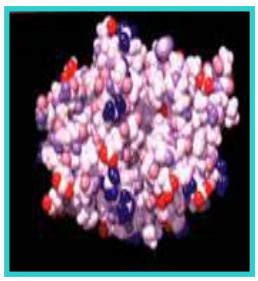


www.integrativebiology.ac.uk

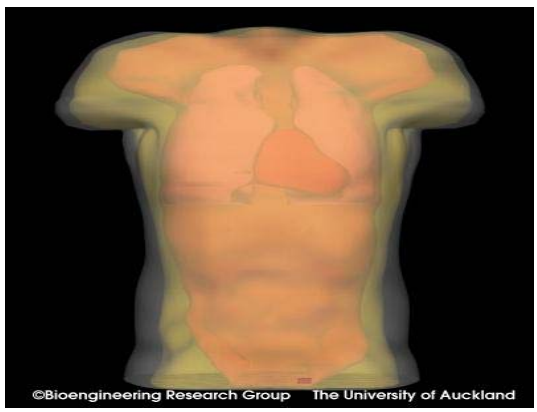
Project Aims

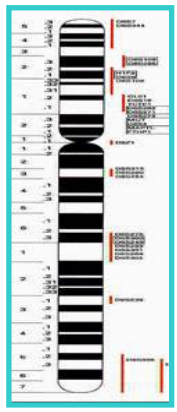
The Integrative Biology Grid will ultimately lead to better drugs for heart disease and cancer - two of the UK's biggest killers. This EPSRC-funded e-Science project which started in February 2004 will build a powerful, fault-tolerant Grid infrastructure for biomedical science which will enable biomedical researchers to use distributed resources such as high-performance computers, databases and visualisation tools to develop complex models of how these killer diseases develop.





Towards predictive in-silico models to guide experiment and, ultimately, the design of novel drugs and treatment regimes





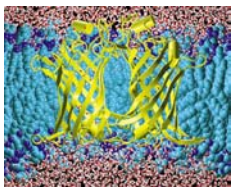
Clinical and physiological researchers in academia and the pharmaceutical and biotechnology sectors will find their work more fruitful through use of these new tools. In addition, the Integrative Biology Grid will benefit the whole UK e-Science community. The toolkit that the project will develop will potentially be useful in other areas which need a total system approach, for example, understanding environmental change processes.

With such tools at their disposal, researchers will be able to model some of the most complex biological systems in the clinical and life sciences. They will be able to "grow" virtual tumours through the crucial stages of early development. Such a process will allow them to uncover the many complex biochemical and physiological processes that take place in tumour growth and potentially spot points at which a drug might be used to stop growth in its tracks.

But most importantly, the ultimate beneficiaries will be patients with heart disease, cancer and other potentially fatal diseases.

Dr David Gavaghan, Associate Director at OeSC and Principal Investigator of the project stresses its importance:

"It is clear that in the post-Genomic era, progress in biological research will depend upon our ability to develop a coherent underpinning theory of biology to allow us to make full use of the vast wealth of experimental data now available. Our aim in this project is to build the e-Science infrastructure to support these endeavours, whilst at the same time investigating the two diseases which together cause over 60% of deaths in the UK - heart disease and cancer."



Participants

The project includes leading biomedical and computing experts from CCLRC and the universities of Oxford, Auckland, Sheffield, Nottingham, Leeds, Birmingham and UCL. This is one of the first projects to come out of the strategic relationship between Oxford University and IBM.

The Integrative Biology Project is an international, multi-institutional, multi-disciplinary project with a particular focus on heart disease and cancer that will demonstrate the crucial role of e-Science in facilitating large scale collaborative research

The University of Auckland · IBM · CCLRC · UCL · EPSRC · The University of Nottingham · The University of Birmingham · University of Leeds · University of Sheffield

The Integrative Biology partners and their roles

Oxford
lead partner, heart, cancer and molecular modelling, HPC, security, project management

CCLRC
Grid services, visualisation, data management, HPC, project architect

Nottingham
cancer tumour modelling

Sheffield
experimental studies of cancer tumours

Birmingham
modelling chemotherapeutics

Leeds
data visualisation, computational steering

UCL
modelling, computational steering, HPC

IBM
data management, Grid data services (OGSA-DAI)

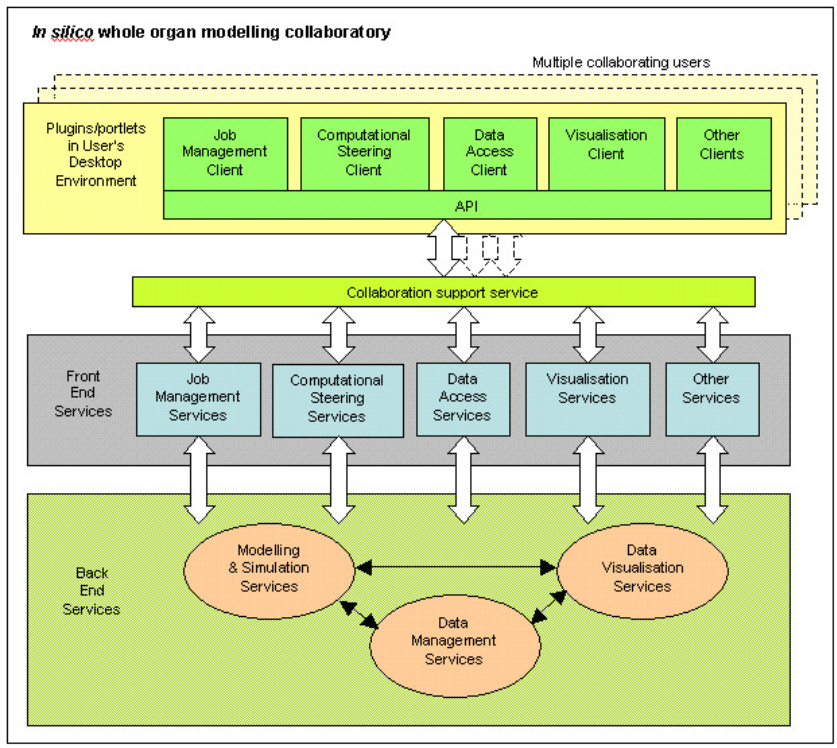
Auckland
expertise and codes for whole organ heart modelling

Software Architecture

Portal/portlet technology, integrated with the user's normal desktop environment, will be used to provide users with a lightweight interface to the operational services developed by the project. These user-accessible services will initially be grouped into four main categories:

- job management (including deployment, co-scheduling and workflow management across heterogeneous resources);
- computational steering (both interactive for simulation monitoring/control and pre-defined for parameter space searching);
- data management (from straightforward data handling and storage of results to location and assimilation of experimental data for model development and validation);
- analysis and visualization (not only of results, but also of interim state, parameter spaces, etc, for steering purposes).

Underpinning development of the architecture are three fundamental considerations: standardization, scalability and security.
Many of the underlying components will be adopted from previous projects, and adapted if necessary in collaboration with their original developers.



e-Science vision and challenges

The e-Science vision driving the Integrative Biology project is of Grid technology enabling the computational biologist to carry out collaborative *in-silico* experiments within a distributed virtual laboratory, or 'collaboratory', which integrates the best available resources for computation, data management, visualisation and data analysis. This environment will facilitate assimilation of clinical and experimental data with computational results to refine models and validate simulations until they achieve the necessary accuracy. This is a second generation e-Science project which will build on many developments emerging from first generation projects and at the same time will itself advance the e-Science agenda on several fronts.

The e-Science challenges for the project include:
- providing transparent, co-scheduled access to appropriate combinations of distributed high performance computers and database resources needed to run coupled multi-scale whole organ simulations;
- exploiting these resources efficiently through application of computational steering, workflow, visualisation and other techniques developed in earlier e-Science projects;
- enabling globally distributed biomedical researchers to collaboratively control, analyse and visualise simulation results in order to progress the scientific agenda of the project;
- maintaining a secure environment for the resources used and information generated by the project without inhibiting scientific collaboration.