**Experimental Physiology – *Review Article***

# Computational physiology and the physiome project

Edmund J. Crampin[1,2,3], Matthew Halstead[2], Peter Hunter[2], Poul Nielsen[2], Denis Noble[3], Nicolas Smith[2] and Merryn Tawhai[2]

[1] *Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24–29 St Giles, Oxford, OX1 3LB, UK*
[2] *Bioengineering Institute, University of Auckland, Private Bag 92019, Auckland, New Zealand*
[3] *University Laboratory of Physiology, Parks Road, Oxford, OX1 3PT, UK*

**Bioengineering analyses of physiological systems use the computational solution of physical conservation laws on anatomically detailed geometric models to understand the physiological function of intact organs in terms of the properties and behaviour of the cells and tissues within the organ. By linking behaviour in a quantitative, mathematically defined sense across multiple scales of biological organization – from proteins to cells, tissues, organs and organ systems – these methods have the potential to link patient-specific knowledge at the two ends of these spatial scales. A genetic profile linked to cardiac ion channel mutations, for example, can be interpreted in relation to body surface ECG measurements via a mathematical model of the heart and torso, which includes the spatial distribution of cardiac ion channels throughout the myocardium and the individual kinetics for each of the approximately 50 types of ion channel, exchanger or pump known to be present in the heart. Similarly, linking molecular defects such as mutations of chloride ion channels in lung epithelial cells to the integrated function of the intact lung requires models that include the detailed anatomy of the lungs, the physics of air flow, blood flow and gas exchange, together with the large deformation mechanics of breathing. Organizing this large body of knowledge into a coherent framework for modelling requires the development of ontologies, markup languages for encoding models, and web-accessible distributed databases. In this article we review the state of the field at all the relevant levels, and the tools that are being developed to tackle such complexity. Integrative physiology is central to the interpretation of genomic and proteomic data, and is becoming a highly quantitative, computer-intensive discipline.**

(Received 21 October 2003; accepted 5 November 2003)
**Corresponding author** P. Hunter: Bioengineering Institute, University of Auckland, Private Bag 92019, Auckland, New Zealand. Email: p.hunter@auckland.ac.nz

Physiology has always been an integrative science concerned with understanding quantitatively how the structure and function of cells, tissues and organs explain the complex behaviour of living systems. During the first half of the 20th century physiologists, from Sherrington and Eccles to Hodgkin and Huxley, revealed the physical basis of human physiology all the way from cell biophysics to integrative control. In the last 50 years, however, the biological limelight has progressively focused on molecular biology, with its spectacular success in explaining mechanisms at the level of genes and proteins. In recent years this success has largely been based on the development of experimental techniques such as DNA sequencing, PCR, microarrays and confocal fluorescent imaging, and above all on the realization that high-throughput measurement coupled to comprehensive databases is just as important to quantitative science as the more traditional approach of hypothesis-driven research. The inevitable consequence of this success is an explosion of data at the subcellular level that are difficult to interpret in relation to the physiological behaviour of complex living organisms. One of the main challenges in physiology over the next 10 years is therefore the interpretation of the genome and ascribing physiological function to genes and proteins in the wider context of integrative systems (Hunter & Borg, 2003; Hunter *et al.* 2002).

The physical sciences, on the other hand, have for the past 200 years confronted nature's complexity with

Authors are listed alphabetically.

 DOI: 10.1113/expphysiol.2003.026740

the development of mathematical models of natural phenomena. Our ability to understand complex fluid flow, for example, in order to design aircraft or forecast weather, is a testament to the physicists and mathematicians of the 19th and 20th centuries, who identified nature's physical conservation laws and developed the mathematical framework to describe them. The successful application of these laws to the solution of engineering problems has also advanced greatly in the past 50 years from the development of computers and numerical analysis [every new complex engineering device is, these days, designed with the aid of mathematical models and finite element (or similar) analysis].

The use of mathematical modelling in physiology gained prominence in the 1950s with the successful prediction by Hodgkin and Huxley of the speed of action potential propagation along a nerve fibre from cable theory coupled to models of ion channel conduction and gating kinetics. Other early successes in applying techniques from the physical sciences to physiological systems were the engineering analysis of blood flow in arteries using computational fluid dynamics, and orthopaedic stress-strain analysis using linear elasticity theory and finite element analysis. In both cases, however, the biological problems are not significantly different from other engineering problems – blood can be treated as a viscous Newtonian fluid like water, albeit with higher viscosity and some unusual characteristics at low shear rates, and bone behaves as an orthotropic linear elastic material, at least until one considers its ability to grow and remodel. It was not possible in these early applications of mathematics and engineering to biology, which pre-date modern molecular biology, to link physiological behaviour to molecular detail.

In this article we argue that the discipline of physiology should now embrace the new era of 'computational physiology' in which mathematicians and bioengineers will work alongside physiologists and molecular biologists to link the physiology of cells, tissues and organs to the growing genomic and proteomic databases. The time is right – genomic and proteomic data are routinely collected; it is clear that the timing and spatial location of gene expression is controlled by environmental factors conveyed to the transcription factors in the gene regulatory regions by complex and redundant pathways which can only be analysed with network models; the physiological significance of genetic diseases will only be understood by linking quantitative models of tissue and organ physiology with the signal transduction cascades, metabolic pathways and other cellular processes. However, the tools required for this so-called 'Physiome Project' [the term 'physiome' comes from 'physio' (life) and 'ome' (as a whole) and is intended to convey a 'quantitative description of physiological dynamics and functional behaviour of the intact organism] are significantly different from the tools of standard engineering analysis for a number of reasons. Biological materials are almost always inhomogeneous, anisotropic and exhibit non-linear behaviour, but nowadays even these characteristics are not unusual in engineering materials. Biological processes (such as signal transduction pathways) exhibit enormous complexity, often with extraordinary degrees of apparent redundancy, but again so do engineered systems such as the electronic circuits in a Boeing 777 aircraft. The really significant and unique characteristic of biological materials is their ability to grow and remodel in response to changing environments – determined partly by genes and partly by their physical environment. An important consequence is that structure and function are intimately linked in a way that no engineering material or system can emulate. Capturing these structure-function relationships in a computationally efficient manner is the key to successful computational physiology and requires models and software that are fundamentally different to those found in the engineering world.

Our 'physiome' theme is developed in several stages. We begin with a discussion of biological complexity and the sources of data for modellers at the molecular level. We briefly describe the framework of ontologies and markup language standards that are being developed for handling our knowledge of biological systems at all levels of biological organization from genes to organisms, integrating across species, age and pathological state. We then discuss progress in modelling biological systems, beginning with gene regulation and cell function and ending with anatomically and biophysically based models at the organ level. The heart and lungs are used to illustrate the use of models that integrate structure and function across multiple spatial scales. Throughout the paper we use the heart as our primary example of multiscale modelling because the heart has provided the first example of a 'physiome model' of an organ.

## Biological complexity

The new millennium began, appropriately enough, with the announcement in February 2001 that the Human Genome Project was close to achieving its goal (International Human Genome Mapping Consortium, 2001; Venter *et al.* 2001). Soon after, in April 2003, the completion of the full sequence was announced (Collins *et al.* 2003). As a result, immense opportunities

for integrative physiology and systems biology have opened up. The challenge to interpret this vast volume of data at the genomic and proteomic levels in terms of function at higher levels is what modern physiology is about, and this is precisely what is meant by integration in biology. It is also, in part at least, what is meant by 'systems biology': the application of systems theory to the complexity of biological interactions at all levels. [This is what distinguishes the word 'system' in 'systems biology' from its long-established use in 'systems physiology'. The latter refers to high levels (such as cardiovascular, respiratory, nervous, immune systems), whereas 'systems biology' can refer to systems theory analysis at any level, including for example gene networks, cell signalling cascades and metabolic pathways].

## Biological complexity and the direction of causality

But why should we bother with such complexity? Having broken biological systems down into their smallest components, have we not begun to reach our goal of understanding, at least in principle? Wasn't the aim to simplify, not to 'complexify'? Could we not now 'compute' the behaviour of living systems (Brenner, 1998) from the automatic and inevitable interactions of their component parts, starting with genes coding for proteins, proteins interacting to form pathways, in turn to form cells, tissues and organs – and so on up to the highest levels? There are at least two reasons why this is not possible. The first reason is that it is computationally impossible and would be lacking in insight (Noble, 2002*a*). Being able to reconstruct a process is not sufficient, by itself, to understand the process. There will be many parameters in the reconstruction equations that will be in need of explanation, and we will also want to achieve more general insight into how the process reacts to perturbations, including most importantly, those of disease states. The second reason is that causation does not simply run in one direction (i.e. upwards). Lower level events, such as gene expression, are controlled by higher level processes. Without understanding the higher level logic in its own right (Noble & Boyd, 1993), the lower level data will often be just that: masses and masses of unexplained data.

## Levels of Selection

But do the higher levels have a logic? Or are they just the 'lumbering robots' (Dawkins, 1976) of gene-level selection, with about as much logic as a cloud formation? The issue of levels of selection is now the subject of important debates in the theory of biology (Williams, 1992; Keller, 1999; Gould, 2002; Krakauer, 2002) The question here is the level at which selection operates. If it really were the case that selection is entirely gene-orientated, then the 'lumbering robots' (i.e. the physiology of higher level function) would indeed be of limited value, even irrelevant to a general theory of biology. However, this gene-orientated view confuses several different processes. Genes are not simply physical stretches of DNA, just as the information on a magnetic surface is not the surface itself. Genes are the carriers of information that any part of an organism can read, information which is transmitted from one generation to another. This transmission is parsimonious in the sense that it is not necessary to code for everything that happens in the development and life of an organism. The properties of water, lipids and the rules of self-assembly, for example, must all be taken for granted. As a 'book of life' the genome is necessarily incomplete (Noble, 2002*b*). It may also be repetitive and redundant, though the extent to which junk DNA is functional, rather than truly redundant, is also an open and controversial issue.

## Replicators and interactors

These considerations are usually expressed by distinguishing between replicators, containing information that gets copied from one generation to another, and interactors, such as macromolecules (even DNA itself, as a chemical substance, is in this sense an interactor), biochemical pathways, cells, organisms, and species and clades, which both carry the replicator information and are subject to selection. Since genes interact so that many genes contribute to any given higher-level function, and each higher-level function depends on many genes for its transmission, natural selection must act at the levels at which functionality appears and this will only very rarely depend on a single gene. One of the challenges for theoretical biology is to identify the levels at which different functions can be said to be expressed and therefore to identify the levels at which natural selection must operate. This is a challenge to which the Physiome Project can respond since, when it is complete enough, it should reveal the modular nature of gene interaction, and the extent of that modularity, at all levels.

## Gene numbers and functions

Analysing biological complexity at various levels is therefore necessary. But there must be very many different

ways of measuring and assessing complexity, some more appropriate to one level rather than another, with each level having its own criteria. One way to look at complexity from a genome level is to ask how many biological functions are available to evolution in any given genome. For a genome of 40 000 genes if we make the extreme assumption that only two genes are needed to define a function then the total number of possible functions would be $0.5 \times 40\,000 \times 39\,999 = 799{,}980{,}000$. Even with this absurdly minimalist view, the number of possible combinations of effects is very large. With more realistic assumptions about the number of genes involved in each function, the figures are really huge: thus at 100 genes/function we obtain $\sim 1.5\ e^{302}$; and if we allow for all possible combinations we get $\sim 2\ e^{166713}$! This has two consequences that are relevant to complexity. The first is that with such huge potential combinations of gene product interactions, trying to understand the system purely at the genome level is clearly impossible. Nature can have used only a minute fraction of these combinations and we need to look at these 'from above' to identify them, i.e. we need to look at the genome through the eyes of higher levels, including physiology. Since only a minute fraction of possible interactions are used, this introduces another way of viewing the question of how contingent evolution has been. With such enormous possibilities, the chances of humans (or any other species) evolving may be exceedingly small. The counterargument would be that evolution really is constrained. But constrained by what? This must be constraint by the logic of life at a higher level. In both cases, though, we need to understand the higher-level logic, either as a constraint or as a chance historical success, or most probably both: chance and necessity interact.

### The (imperfect) logic of life

We are a long way from being able to judge such broad questions. But it is worth noting here that we would not expect the 'logic' to be perfect from an engineering point of view. Evolution is a dynamic process and there is no reason to suppose that what has evolved is fully in equilibrium with its environment (though Gould's view of stasis in evolution comes close to this – see Williams, 1992, chapter 9). Moreover, evolution is an historical process, with all the accidents and dead ends that a blind journey will produce (see particularly chapter 6, 'Historicity and Constraint', in Williams, 1992). Genes are perhaps better viewed as prisoners of the successful physiological systems that carry them than the determinants of those systems. They simply, and necessarily, code what has worked well at the level of the interactors. 'Worked *well*' is the correct description. 'Works *best*' would require design that avoids the historical accidents (using existing structures for a new function, such as the formation of the mammalian middle ear from modified jaw and hyoid arch bones) and evolutionary traps (such as the convoluted path of the sperm ducts following descent of the testes, or the inversion of the retina). Ultimately, we must be able to account for the evolutionary history of functional development. As Jared Diamond has insisted (Diamond, 1993), physiology should re-connect with the mainstream biological theory of evolution.

### Comparison of genomes

These calculations put into perspective the frequent comparisons of similarity of genomes. If we differ from species x by only 1% then, from a genome point of view, the differences look insignificant. Using the same calculations as above, however, we get a totally different perspective. Introducing less than 1% of 1% of a genome of 40,000, i.e. a *single* gene, *doubles* the available functions. The argument is remarkably simple. We still have all the possible functions of the original set of genes, and all of those again in combination with the new gene. If we differ from another species at the genome level by, say, 1% (i.e. 400 genes) then the potential differences at a functional level are truly enormous. Even a 0.1% difference corresponds to 40 genes and also a huge number of new possibilities. To these calculations, we have also to add the influence of the environment through its control of gene expression levels. Suppose, for example, that gene expression levels might vary for a given protein by around an order of magnitude. This would mean that, to the original 40 000, we can add a further 400 000 elements to play a part in combinations. That puts the nature–nurture argument in perspective, and the absurdity of trying to arrive at percentages for the influence of each. Of course, only a very small proportion of the possible combinations actually occur.

## Functional genomics and molecular physiology

As an example of such a comparison of genomes, how can the apparent complexity of the human body be understood in relation to a much smaller and simpler organism, for example the nematode worm *C. elegans*, which appears to have more than half the number of protein-coding genes but on the order of $10^{11}$ fewer cells. [Each of the cell divisions taking place during development of the worm from fertilized zygote to mature adult has been described in exquisite detail, giving the complete lineage for the 1000 or so cells making up the adult

worm (www.wormbase.org). While the relatively simple physiology of the worm can be understood in terms of the different cell types and their interactions, it is still far from understood in terms of its genome]. If the gene numbers predicted by the various genome projects prove to be accurate, then the data confirm the view that it is the enumeration of possible interactions between genes, rather than the absolute size of the genome, that determines biological complexity. It is therefore the regulation of these interactions which must be the major postgenomic focus (Claverie, 2001). [Although the number of bases in the human genome is determined, there is still much dispute over the number of protein-coding genes, and proposed gene numbers are estimates that are sensitive to the gene-finding algorithms used. However, the waters are further muddied by the observation of highly conserved sequences in the more than 99% of the genome which apparently does not code for proteins. Furthermore, homology between species is found in these non-coding regions suggesting a functional, possibly regulatory role for the so-called junk DNA (Dennis, 2002; Mattick, 2001)].

The enormous challenge lying ahead of the human genome project, and the ultimate goal of functional genomics, is to understand the role of the genome in development, normal physiology and, perhaps most significantly, in disease. New experimental technologies are accelerating the identification of patterns of gene expression that delineate tissue types and underlie pathology. Altered gene expression in response to changing stimuli or through disease gives rise to reprogramming and changing function within tissue. Examples in cardiac ventricular tissue that have received recent attention include the molecular changes underlying the sequence of structural and functional modifications during remodelling after myocardial infarction (Cohn *et al.* 2000; Sehl *et al.* 2000; Stanton *et al.* 2000) and those responsible for modified electrical properties generating cardiac arrhythmia (Antzelevitch *et al.* 2001; Noble, 2002*c*). Variation in gene expression between individuals gives rise to increased disease susceptibility or responsiveness to therapy and it is a major aim of functional genomics to unravel the combinations of environmental and genetic factors that underlie complex disease, which to this point have proved elusive. These examples highlight the position of functional genomics at the interface between molecular biology and physiology. [This has been recognized in recent titles for the quadrennial IUPS International Congress of Physiological Sciences, held in 2001 in Christchurch, New Zealand, under the banner 'from Molecule to Malady', and due to be held in San Diego in 2005 with the focus 'from Genomes to Functions].

Progress on understanding the role of the genome in cell, tissue and organ function is a highly integrative problem, which requires significant computational input. The major emphasis is currently on genome (and proteome) informatics: databasing, annotation and so forth (Winslow & Boguski, 2003). In addition to sequencing and annotation projects for different species, genomic databases with a particular tissue or disease focus are now being developed, for example, the Cardiac Gene Expression knowledgebase (CaGE; www.cage.wbmei.jhu.edu), which aims to catalogue all genes expressed in human cardiac tissue (Bober *et al.* 2002). While this is of great importance for making genomic data available and usable, it is increasingly apparent that the challenge ahead is to give quantitative predictions to physiological outcomes from genomic knowledge. As more and more data accumulates, a computational framework is required which can reach down to the genomic level via modelling of regulatory processes; in particular gene regulatory networks, interactions with signalling and metabolic pathways and how these influence and effect higher level organization and function.

The trend in current biology is towards comprehensive, automated measurement techniques. The development of high-throughput technologies for the systematic measurement of gene transcripts, proteins and their interactions, coupled to advances in computational data processing, provides a method for probing the functional role of large numbers of genes. DNA microarrays have emerged as the pre-eminent technology for large scale parallel gene expression studies (see, for example, The Chipping Forecast II, 2002).

### Gene expression profiling

Microarrays exploit the complementary binding properties of nucleic acids: fluorescently labelled gene transcripts (mRNA) isolated from a tissue sample are allowed to hybridize with molecules of known sequence (ESTs or cDNA) immobilized at particular locations in an array. The fluorescence intensity at a particular spot on the microarray reveals the amount of the gene transcript which has hybridized at that location and the ratio of intensities of red and green fluorescent labels on the same array provides a measure of the differential expression level of the corresponding genes between two samples. These data on the up- and down-regulation of genes can be collected across different tissue types, disease

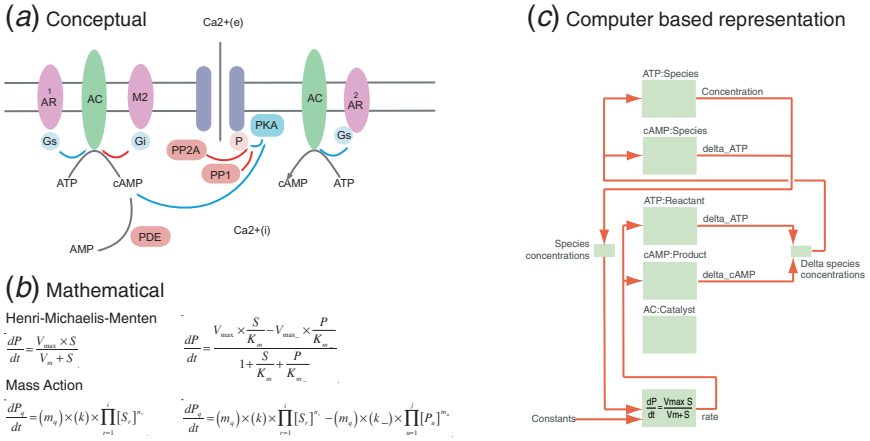


**Figure 1. Model representations**
*A,* conceptual: representation of the model through diagrams, natural language, or concepts from ontologies. *B,* mathematical: representation of the conceptual model using mathematical equations. *C,* instances: representation of mathematical equations as instances of CellML components connected together.

states and in response to different stimuli. Data collected on temporal gene expression profiles, by sampling mRNA at several different times following experimental perturbation, provide gene expression time series data which are crucial for the development of dynamic models of gene regulation.

Studies of human disease using microarrays for gene expression are increasingly common, as shown in several recent reviews on microarrays in cardiovascular research and medicine (Cook & Rosenzweig, 2002; Napoli *et al.* 2003). Disease and tissue-specific microarrays have been developed – for example, a cDNA 'CardioChip', comprising over 10 000 cardiovascular-related ESTs (Barrans *et al.* 2001; Barrans *et al.* 2002). Most of these studies aim to identify differentially expressed genes (Henriksen & Kotelevtsev, 2002; Simkhovich *et al.* 2003). Clustering gene expression profiles is a common starting point for the analysis of microarray data and is used to find functional groupings of genes (Eisen *et al.* 1998; Slonim, 2002). If the expression profiles of a set of genes are closely correlated over many different experimental conditions then this co-expression is taken as a possible indication of a similar functional role for the cluster of genes or, even, of shared regulatory control mechanisms. Similarly, principal component analysis (PCA) can be used to reduce the data to a few simple underlying expression patterns, in particular for periodic temporal patterns of gene expression (Alter *et al.* 2000; Holter *et al.* 2000). The 'guilt–by association' of new genes with genes of known function via clustering suggests new targets for study.

## Data requirements for modelling regulatory networks

High-throughput gene expression techniques generate large amounts of data. The primary concern for the analysis and modelling of large scale data sets is the so -called 'curse of dimensionality': the parameter space for a model grows exponentially with the number of variables (genes). This makes finding appropriate parameter values for large scale gene network studies a major challenge. The key to this problem is the use of prior knowledge on interactions and components of the network which have already been identified, features common to biological networks such as sparseness (Jeong *et al.* 2000), and the use of simplified response functions in the model, can all be used to reduce the number of undetermined parameters. To fully characterize network behaviour it is necessary to sample gene expression profiles under many different combinations of inputs and perturbations,

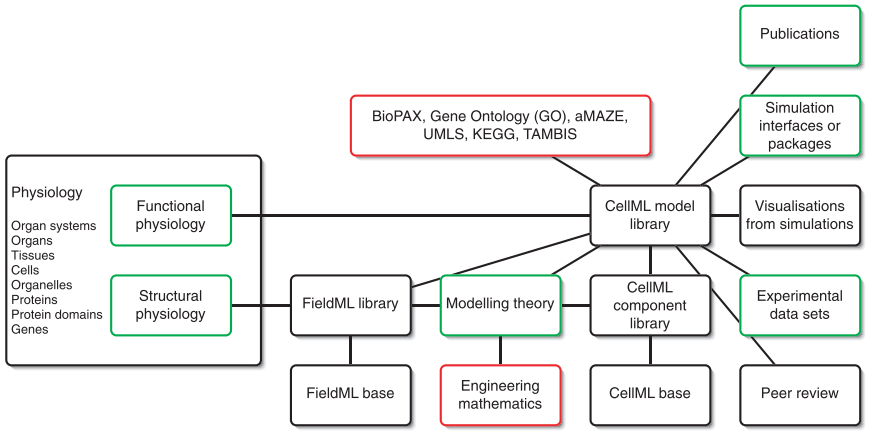


**Figure 2. Ontology domains in the physiome project**
Black boxes, ontology domains that the IUPS project needs to develop; green boxes, ontology domains in which the IUPS ontologies will share a common framework with others; red boxes, ontology domains that the IUPS project does not need to explicitly build.

for example point mutations, gene deletions and over-expression, inhibition of translation by antisense RNA, and perturbation of signalling and metabolic pathways.

Localized information on gene expression, required for spatially extended models, can be obtained from immunofluorescence studies. Laser capture microdissection (LCM), which automates the dissection of specific cells and multicell structures which have first been identified under the microscope (Emmert-Buck *et al.* 1996; Mills *et al.* 2001), allows gene (and protein) profiling studies to be conducted at high spatial resolution. The quantity of mRNA which can be extracted from such small samples may not be sufficient for high quality microarray studies. For detection of low abundance transcripts quantitative reverse-transcription PCR (RT-PCR) can give high sensitivity measurements of transcript levels, with the added advantage that RT-PCR can measure absolute, rather than relative concentrations of RNA (Schnell & Mendoza, 1997). PCR primers are required for each gene of interest and so, as with microarrays, the technique cannot be used to analyse previously unidentified genes. Transcripts of unknown genes can be measured using Serial Analysis of Gene Expression (SAGE; Velculescu *et al.* 1995; Yamamoto *et al.* 2001; Patino *et al.* 2003), where large numbers of transcripts are counted and analysed efficiently by sequencing only short tags from each RNA molecule. These short sequences can usually identify the corresponding gene (as long as its sequence is known). All tags collected from a sample are assembled into a single molecule which is then amplified and sequenced, providing quantitative information on the amount of each transcript in the sample.

## Proteomics and metabolomics

Gene expression profiles describe the transcriptional state of cells or tissue, but do not directly reflect protein abundance. Proteomics and metabolomics aim to provide an unbiased identification and quantification of all proteins and metabolites present in a biological sample. The proteome encompasses modifications of protein molecules and interactions between proteins. Knowledge of the changing protein content of cells, and functional interactions between proteins is of primary importance in developing models of cellular function. Gene expression studies provide the link between the (relatively) static genome and the highly dynamic protein content of a cell, responding to environmental changes, signalling pathways, as well as normal homeostatic protein turnover. The relation between gene expression profiles (the transcriptional state), as measured by mRNA level, and net protein synthesis depends on translation and post-transcriptional modification rates, as well as on the rate of protein decay, whether the molecule is actively degraded or transformed.

Protein molecules are intrinsically less amenable to high-throughput studies than are nucleic acids due to their secondary and tertiary structure. Technologies for large scale proteomics analyses include separation techniques (2D-PAGE, Poly Acrylamide Gel Electrophoresis) and mass spectrometry. Microarray-based techniques are also available for protein studies, although these platforms are less developed than DNA microarrays, both for detecting protein abundance by using different affinity reagents (for example antibodies, as a multiplexed immunoassay) to bind to proteins selectively, and for protein–protein interactions (MacBeath, 2002), also studied in the Yeast Two-Hybrid system. The identification and measurement of interactions, both stable and transient, between biomolecules *in vivo* is probably the widest current knowledge gap, and greatest experimental challenge.

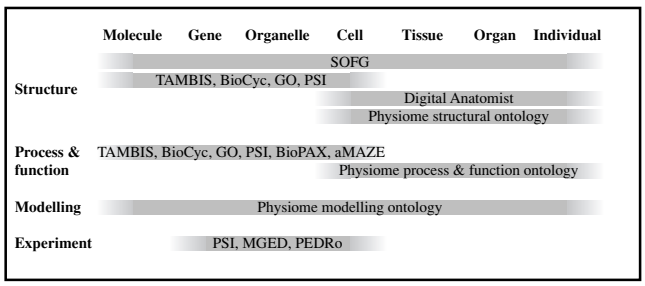


**Figure 3. Ontological domains**
SOFG, Standards and Ontologies for Functional Genomics (www.sofg.org); GO, Gene Ontology (www.geneontology.org); MGED, Microarray Gene Expression Data (mged.sourceforge.net); BioPAX, Biological Pathways Exchange (www.biopax.org).

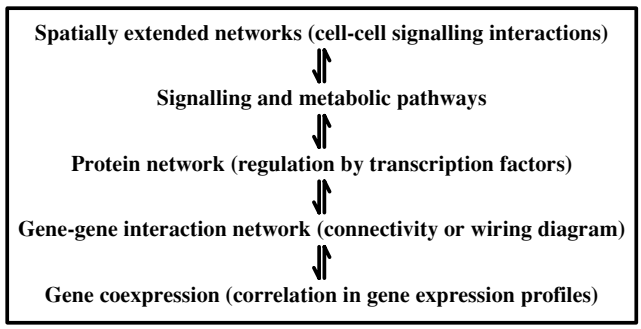


**Figure 4. Hierarchy of network models for gene regulatory network modelling**

### Issues and standards for gene expression profiling experiments

Important for the development of robust and predictive models is the quality of the data available, good data annotation, and a methodology for assessing and managing the sources of variability in the data. There remain many issues to do with data reproducibility, variability and experimental design for microarray experiments (Churchill, 2002; Speed, 2003), which we will not discuss here. Differences between nominally similar tissue samples reflecting normal biological variability need to be quantified to minimize false positives in the identification of differentially expressed genes. Modelling may be a useful tool for identifying the sources of robustness to normal biological variability in regulatory networks (Alon *et al.* 1999; von Dassow *et al.* 2000).

The details of experiments performed, on sample preparation and so on, have significant bearing on the interpretation of profiling studies. With this in mind, ontologies and languages for the annotation of gene expression and proteome data are being developed (Stoeckert *et al.* 2002; Taylor *et al.* 2003). Standards have been proposed for describing microarray experiments, MIAME (Minimum Information About a Microarray Experiment) by the Microarray Gene Expression Data society (MGED; www.mged.org), which in late 2002 were adopted by several major journals including Nature, The Lancet and Cell as requirements for publication of the results of microarray studies. Annotated data is required to be deposited in one of several public databases: for example Gene Expression Omnibus (GEO) at NCBI (www.ncbi.nlm.nih.gov) and ArrayExpress at EBI (www.ebi.ac.uk/arrayexpress), which conform to the MIAME standards. A mark-up language for data exchange between databases, MicroArray Gene Expression Markup Language (MAGE-ML; www.mged.sourceforge.net), has been developed by MGED and others, under the auspices of the Object Management Group (OMG; www.omg.org).

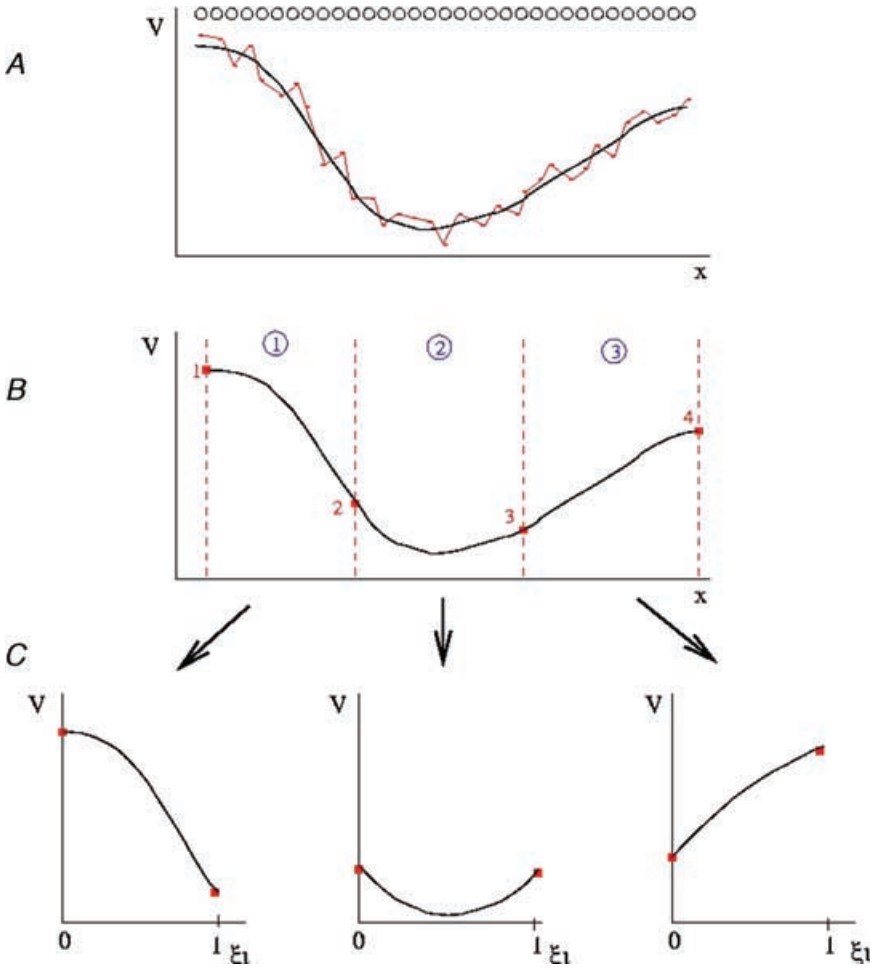


**Figure 5. Finite element representation of cellular potentials**
*A,* a one-dimensional row of myocytes, each with transmembrane potential shown below by the dots. *B,* the mean potential distribution is approximated by a smooth curve constructed from piecewise polynomials. The three sections separated by dashed lines are called *finite elements. C,* the field inside each element is given by an interpolation of potentials defined at the nodes (shown by the red dots).

## Physiome ontologies and markup language standards

To link the rapidly growing knowledgebase of biological data into a physiome modelling framework, formal vocabularies need to be developed to reduce the growing heterogeneity of terms. This is especially important as models of physiological processes are developed that span multiple spatial scales (genes and proteins to cells, tissues and organs) and incorporate the parameter changes associated with disease. A formal representation of custom data structures in the many applicationspecific databases is also needed to provide a common interface between them. Standards must be developed to formalize the description of both experimental data and mathematical models of physiological processes. Ontologies that incorporate semantic descriptions of modelling concepts make the modelling environment richer and unambiguous. They also permit the integration of data from ontologies and databases in other areas of biological research and the building of software tools that interpret and use this information. [An 'ontology' is a taxonomy together with a set of domain-specific rules for linking objects within the taxonomy. For example, a taxonomy for the human musculo-skeletal system names all the bones, muscles, tendons and ligaments, etc. in the human body. The rules are typically biological constraints such as 'the semimembranosus muscle originates from the ischial tuberosity and inserts into the posterior part of the medial condyle of the tibia'].

### Levels of representation

The physiome modelling framework can be considered at three levels:

**1** *Biological* – representation of the biological problem using terms from the biological domain

**2** *Mathematical* – mathematical formulation of the biological problem

**3** *Computer based representation languages* – representations of models in a formal language that is machine interpretable, and which represents computational abstractions of entities, mathematical relationships and rules for their interpretation.

The goal is to represent the details of these levels in a way that can be used to explore the meaning of ideas and observations across the levels. Biology is usually represented using biological terms and relationships in a natural language context. Mathematics is a language in which we can represent theories for the processes and structures that the biological terms and relationships describe. At the computer representation language level, both the biological and mathematical interpretations need to be represented in constructs that can be used in various computer applications, such as simulation and visualization tools, and data repositories.

The three levels of representation are intimately related. Biological concepts link to mathematical concepts and simulation architectures, which link to computer based representations. At each level there is a description of the model, and at each level we find information that feeds across into the other levels. An example; a biological concept of predator–prey interactions can be modelled mathematically as two coupled ordinary differential equations. The way we express this model makes a certain claim on the relationship of the rates of change of predators and prey. Simulation of the model helps us to understand the longterm behaviour of such a description, and allows us to measure data appropriate for validating the biological reality of the model.

The requirements for this multilevelled modelling framework to be effective are: (i) tools for experts to represent models and solutions within particular levels using the language and interfaces that are natural to them, and (ii) computer based representations of what is described using these tools.

One toolset under the first of these requirements that is being developed is a visual editor. Visual building blocks and interfaces are a very natural way for people to navigate libraries and to build and interpret models (see Fig. 1). To develop a framework for building toolsets, we need to have good solutions to the second requirement, i.e. computer based representations of what modellers are describing. To this end, various representation languages, including ontologies are being developed. The integration of ontologies into the physiome project will provide an unambiguous, and machine interpretable, representation of concepts across these levels of modelling, helping to communicate biological models through tools for building, sharing, interpreting, and visualization. One such representation language, CellML (www.cellml.org) has been developed over the last four years, and aims to represent models at the cellular level. The CellML language itself is a set of constructs that have elegant interpretations within both the computational simulation domain and the object orientated programming domain. As a modelling representation language, it is sufficiently generic to represent any mathematical representation of any biological model, not just cells, so serves as a base for the development of a more generalized modelling representation environment. The

evolution of this generalized representation environment includes the integration of ontology data, which provides a machine-interpretable pathway through the levels of modelling, and the further development and integration of FieldML (www.bioeng.auckland.ac.nz/physiome/markup.php) which facilitates the representation of structural and continuum based information about biological and physical entities.

### Building and integrating ontologies

Ontologies are a vehicle for providing unambiguous descriptions of terms and their relationship to one another. To a computer scientist, they provide a formal framework for describing the properties and relationships of concepts that have both a formal logical foundation and a structure amenable to machine processing, interpretation and sharing. To a biologist or modeller, ontologies provide a thesaurus and structure for understanding and binding terms, ideas, data sets, and visualizations, etc.

Many different groups are constructing ontologies for various biological domains. One approach to integrating ontologies from different biological domains would be to generate a large composite ontology. However, this is not the intention of the IUPS project's ontological framework because the biological ontologies that currently exist do not form pieces of the same puzzle – they may have biology in common, but it usually stops there. There is no currently agreed core framework or methodology that can be used to guide the development of compatible domain specific ontologies, but there are efforts underway to promote such a platform. The Unified Medical Language System (UMLS), for example, is attempting to bring together various ontologies from different domains into a composite ontology that fully integrates these knowledgebases.

The current view on the IUPS project's ontological framework is shown in Fig. 2. Some new ontologies are being built from scratch while some existing ones will need to be extended and integrated as both a common framework and data source (i.e. a composite approach). The focus at present is to describe constructs for interpreting our computer based model representations within the biological and mathematical

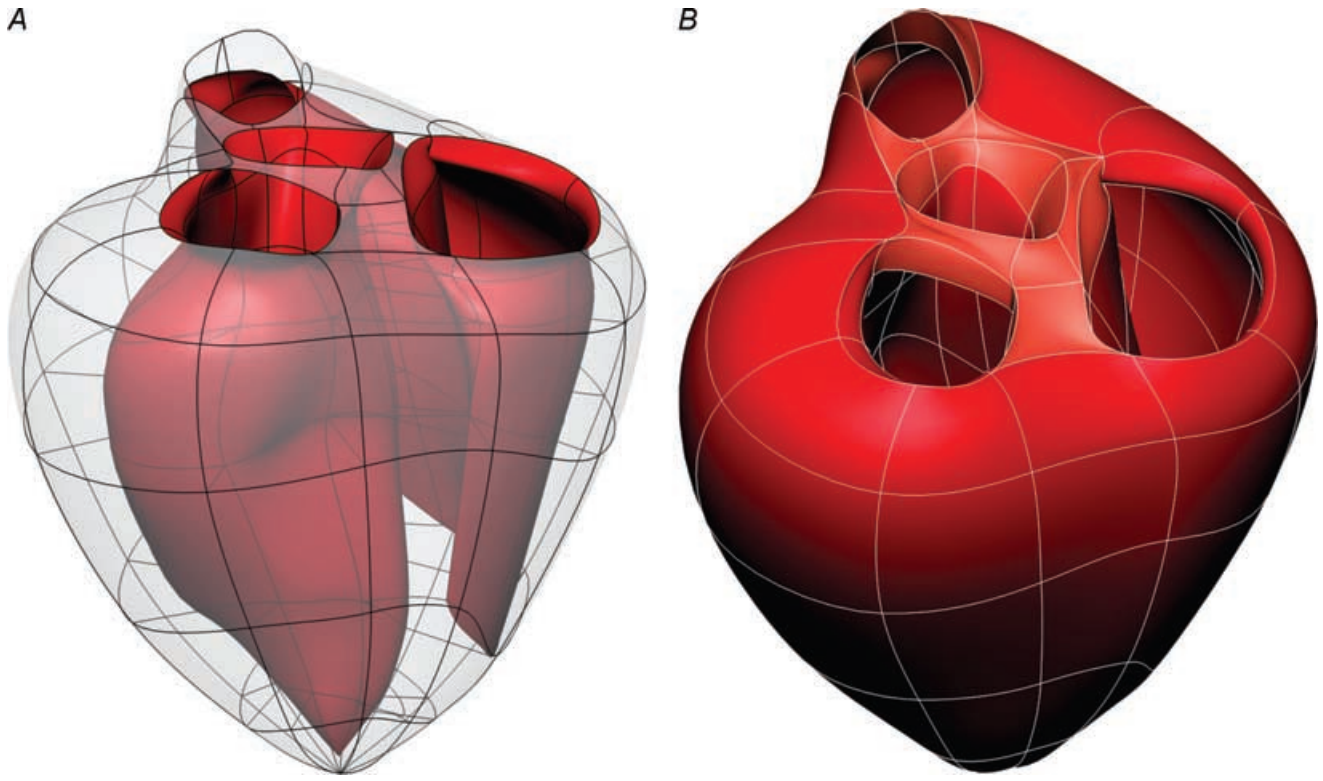


**Figure 6. Geometric models of the heart**
*A,* finite element surfaces fitted to measurements from the left and right ventricles of the pig heart. *B,* 3-D finite element model of the heart. The elements use high order basis functions (cubic Hermite) and therefore relatively few are required to provide an accurate description of ventricular anatomy. From Stevens & Hunter (2003) with permission.

levels of modelling. The domains of modelling theory, and the ML library domains (Fig. 2) capture representations of mathematical relationships, model architectures, and component structure (both physical and abstract).

Ontologies within the data, simulation, and visualization domains provide a top level interface to the resources they describe. The hierarchy of modelling shown earlier describes levels at which a modeller thinks about biological terms, for example – a particular organ or cell, a particular process such as ion transport. These processes and entities are concepts within domains of biology that already have databases and associated ontologies. Instead of defining one particular interpretation of these concepts we can use these other ontologies directly to describe any biological aspect we are referring to in a particular model or ontological concept in our domain. One of the necessary aims of groups such as TAMBIS (Transparent Access To Multiple Bioinformatics Information Stores;

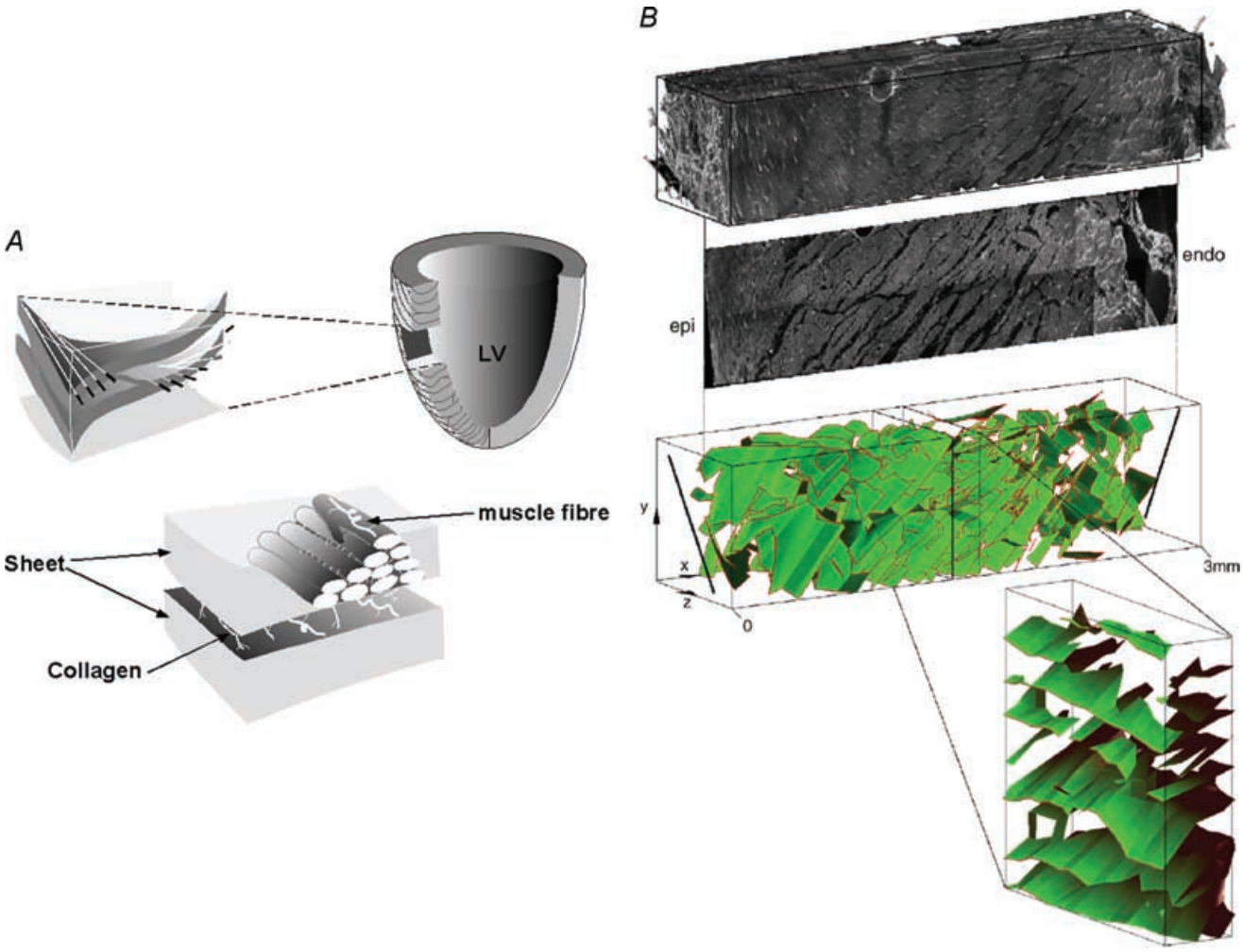


**Figure 7. The microstructure of cardiac tissue**
*A*, schematic diagram showing the variation in muscle fibre direction across the wall in a segment removed from the left ventricle (top), and the branching laminar structure of myocardium in which the sheets are composed of myocytes bound in layers, 3–4 cells thick, by endomysial collagen and surrounded by perimysial collagen which also links to the adjacent sheet (bottom). This 'fibrous-sheet' architecture allows for shearing to occur between the layers and aids the process of wall thickening at end-systole. *B*, a composite 3D confocal image of a transmural block of myocardial tissue from the rat heart with a single image slice shown below (top), and a finite element model of the cleavage planes from the transmural tissue block used to study the flow of current and propagation of the tissue activation wavefront around the myocardial sheet structures (bottom). The spatial variation of tissue properties, such as the density of collagen, gap junctions and ion channels, etc. is defined by the markup language FieldML (www.bioeng.auckland.ac.nz/physiome/physiome.php). Figs *A* and *B,* from Hooks *et al.* (2002) with permission.

imgproj.cs.man.ac.uk/tambis), BioPAX (Biological Pathways Exchange; www.biopax.org) and PSI (Protein Standards Initiative) is that they work together to ensure that their biological concepts are compatible (Fig. 3). In the area of biochemical pathways, the CellML developers are working closely with BioPAX and SBML (Systems Biology Markup Language; www.sbml.org) to establish the foundations for binding cellular domains. An example use-case of such a binding is a pathway of inference starting at concepts in the BioPAX database and ending in selections of models from the CellML database. A number of use-cases are described in the Appendix.

## Modelling gene function

A major effort is underway to process and interpret microarray data to infer regulatory mechanisms in gene networks. Predictive mathematical models of gene networks provide a quantitative framework within which gene expression data can be used to determine regulatory interactions between genes, and the effect on cellular function. Traditionally this approach has been tackled in an intensive fashion, by piecing together available information on individual gene interactions to reconstruct network behaviour. The new data collection technologies demand a change of emphasis to automated data analysis and modelling procedures. The task, to determine the underlying network, its structure and dynamics, from the data has become known as reverse engineering of gene regulatory networks. Many different modelling approaches have been applied to studying gene networks, reflecting differing levels of description and differences in data available to the modeller. A hierarchy of modelling levels is presented in Fig. 4. In addition to these different levels of description of regulatory mechanisms, there are many possible mathematical representations available to the modeller. The choice of mathematical modelling framework will reflect the available data, and also the level of description that is required, from discrete stochastic models which represent individual molecular dynamics to deterministic spatially extended continuum modelling of gene expression.

### Boolean networks

One way to greatly simplify the mathematical representation of networks is to ignore the details of molecular interactions and focus instead on their outcomes, namely whether a gene is 'on' or 'off', according to whether its transcription level is above a given threshold. Regulation is then represented by logical operations (AND, NOR, etc.) on the gene expression states, according to whether interactions activate or repress transcription. Gene expression levels are logical variables which are updated synchronously according to a rule table (a set of if then instructions) which describes the logical operations representing the regulatory interactions between genes. A Boolean network thus represents a 'wiring diagram' for the gene network (Bolouri & Davidson, 2002; Davidson *et al.* 2002). The aim of a reverse engineering approach is to infer the logical rule table directly from data. Techniques have been proposed which demonstrate that in principle a Boolean network can be constructed from data using no prior knowledge (Liang *et al.* 1998).

Although the properties of Boolean networks are much simpler than their continuous variable counterparts,

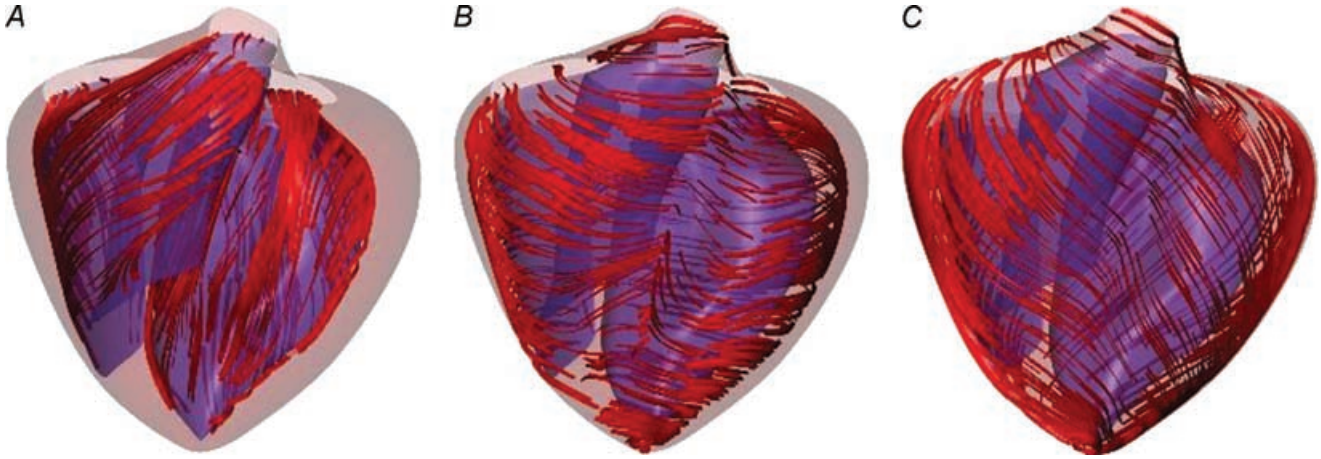


**Figure 8. Fibre orientation in the heart described by 3D finite element fields**
Fibres are shown on *A,* endocardial surface, *B,* midwall, and *C,* epicardial surface. From Stevens & Hunter (2003) with permission.

they retain many of the properties of networks that are important in terms of gene function. For example, steady states are achieved as the network settles down into a stationary state or a repeating pattern of states (oscillation). The appeal of this modelling framework is its simplicity, but much biological detail is clearly lost. In reality, gene expression levels recorded in time course microarray studies spend much of their time at 'intermediate' levels, rather than quickly saturating at maximal expression rate, or falling to negligible levels. Logical networks may therefore be a good modelling strategy when the data quality is poor, and where intermediate expression levels cannot be resolved.

### Kinetic modelling

Systems of differential equations provide a very natural modelling framework for the kinetic behaviour of gene networks, which easily extends to encorporate stochastic effects on gene transcription (McAdams & Arkin, 1999; Kepler & Elston, 2001) and modelling of transport processes in spatially extended systems. For large networks, the 'curse of dimensionality' limits the possibility of inferring parameters of a model from data. A reverse engineering technique has been proposed for linear systems, motivated by the observation that gene networks are sparse (there are relatively very few regulatory interactions between genes) demonstrating that progress can be made even with limited data (Yeung *et al.* 2002).

For well characterized networks, in which the relevant genes have been identified and the wiring diagram determined, biophysically realistic nonlinear kinetic functions can be assumed for the regulatory interactions. This allows quantitative prediction of transcription rates etc. in response to perturbations of the network. Kinetic parameters can be determined for individual reactions (for example the binding of transcription factors, degradation rates of mRNAs, etc.) using time-course data from microarrays, or GFP reporter gene approaches (Ronen *et al.* 2002). Interactions between cells are critical in determining ptterns of differentiation into distinct tissue types. Models of spatially extended systems must therefore incorporate mechanisms for signalling between cells, transport of gene products as signalling molecules and signalling networks within cells (Mjolsness *et al.* 1991; von Dassow *et al.* 2000; Davidson *et al.* 2002).

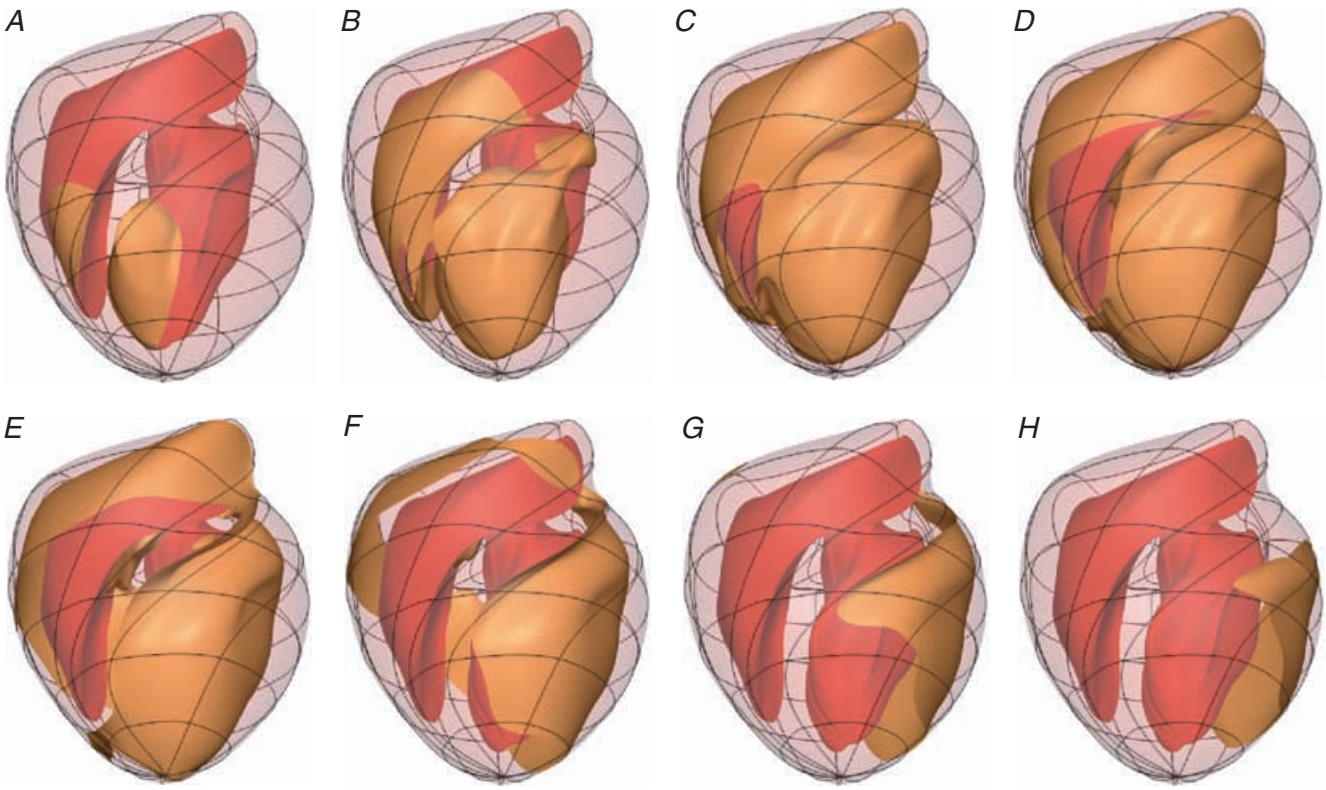


**Figure 9. Wave propagation in the heart**
Wavefront locations using an eikonal equation to simulate propagation from the distal ends of the Purkinje tree. For each sample time anterior (top) and posterior (bottom) views are given. *A* to *H* is 5 ms to 40 ms in 5 ms increments. The endocardial surfaces of the right and left ventricles can be seen through the translucent outer epicardial surface. From Tomlinson *et al.* (2002) with permission.

### Cellular modelling

The coupling of gene expression to tissue and whole organ function requires a number of intermediate models at physiologically based spatial scales. A number of groups have developed sophisticated representations of ion transport between subcellular compartments and via the membrane in, for example, cardiac (Noble *et al.* 1998 and Luo & Rudy, 1994), pulmonary (Smirnov & Aaronson, 1994), and smooth muscle (Yang *et al.* 2003) cell types. Based on systems of coupled ordinary differential equations, the development of these modelling techniques is illustrated below for the cardiac myocyte by considering their application to understanding excitation, contraction and metabolism.

## Multi-scale modelling in the heart

Modelling provides a quantitative framework for establishing the effect of molecular changes on the function of cells, tissues and whole organs. For the heart, modelling is already at an advanced stage: at the cellular level, where electrophysiological, excitation-contraction coupling, contractile and metabolic processes have been incorporated; at the tissue level, where structural details and conduction properties are included; to models describing the anatomy of the heart itself. Models of cardiac myocytes are now sufficiently sophisticated that the effects of up and down regulation of specific genes can in some cases be predicted. These models can characterize disease at the cellular level (e.g. altered protein expression levels or mutated ion channel function), which may then be studied in tissue and whole organ models to predict outcomes on whole organ function. (Hunter *et al.* 2001; Noble, 2002*d*).

### Cardiac cell models

The key cellular processes which are currently characterized within the organ level modelling framework are ion channel electrophysiology, myofilament mechanics and cellular metabolism. The most advanced of these models are the ion channel models pioneered by Noble (DiFrancesco & Noble, 1985; Noble *et al.* 1998; Noble & Rudy, 2001) following the seminal modelling work of Hodgkin & Huxley (1952) on squid axon excitation. The Luo & Rudy (1991, 1994) models are alternative frameworks focusing on the physiological behaviours of premature stimulation and arrhythmogenic activity of the single myocyte. Their modelling framework has since been extended by Jafri *et al.* (1998), among others, to accommodate more complex calcium kinetics which are important for contraction coupling.

In parallel with electrophysiology, Wong (1972), Panerai (1980), and Guccione & McCulloch (1993) have developed specific cardiac models of active tension generation largely based on experimental data from inhomogeneous papillary preparations. Hunter *et al.* (1998) has published a model fitted from more current data measured from intact and skinned trabeculae and the isolated cardiac cell. This model consists of two components, the binding kinetics that regulate the maximum number of cross-bridges (the force-generating bonds formed between contractile proteins) and the phenomenological characterization of the binding and unbinding kinetics of the cross-bridges themselves.

The modelling of metabolic regulation in cardiac cells has, until recently, lagged behind the characterization of excitation and contraction. This is in part due to the immense complexity of what are highly integrated

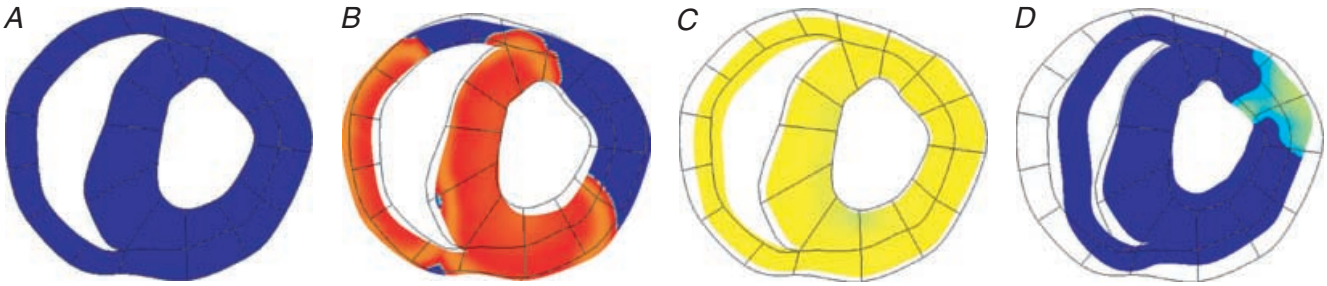


**Figure 10. An anatomically accurate two dimensional model of coupled excitation-contraction in the cardiac ventricles**
The wave of cellular *trans*-membrane voltage (scaled between – 85 mV blue and + 45 mV red) is shown on the deforming tissue model. The undeformed finite element mesh to shown to provide a reference for tissue deformation.

metabolic pathways which exhibit multistate nonlinear regulation via a large number of state variables (Smith *et al.* 2002). However, a number of recent studies focusing on regulation of ion transport (Smirnov & Aaronson, 1994; Michailova & McCulloch, 2001; Ushimaru & Fukushima, 2002) and contraction (Saucerman *et al.* 2003; Smith, 2003) in the myocyte indicate that a fully coupled excitation-contraction model regulated by metabolism will soon be developed.

These cellular-based models now encompass sufficient biophysical detail such that the mediation of specific membrane exchanger and pump ionic currents via gene expression levels can be directly accounted for and linked to specific pathologies (Snyders & Chaudhary, 1996; Clancy & Rudy, 2002; Winslow *et al.* 1998; Antzelevitch *et al.* 2001; Noble, 2002*e*; Mazhari *et al.* 2001). However, to conclusively link gene regulation with clinically relevant whole organ pathology requires a further step in the increasing spatial scale. A continuum framework supplies an effective means of representing the spatial properties of geometry, conductivity and stiffness tensors, and the spatial variation of different cell types and properties. As will be demonstrated in the following section, using cellular models of active function at grid points embedded in the continuum geometry, in combination with governing equations, and appropriate numerical solution methods, presents a powerful tool for linking function through a number of spatial scales.

## Tissue modelling

The major physical processes operating in the heart at the level of the intact organ are (i) the mechanical deformation of the myocardium (ii) the fluid mechanics of blood flow in the atria, ventricles and coronaries (iii) electrical activation of the conducting system from the sino-atrial node to Purkinje fibres and the myocardium, and (iv) transport of metabolic substrates between the coronaries and the myocardium to support the energy demands of the working heart. [Heat flow associated with temperature gradients does not appear to be significant in the in -vivo heart – unlike the lungs where warm blood is in close proximity to cold air]. These tissue level processes are of course supported by a diverse range of physical processes at the cellular level, as discussed in earlier sections. In each case the physical process (mechanics, electrical current flow, etc.) is governed by well-established conservation laws. In the following sections we describe these laws and how they are linked to physical processes at the cell level.

## Continuum fields

A key concept underlying all of these tissue-level equations is the concept of a *continuum* representation of spatially varying quantities. This is illustrated for the spatial variation of transmembrane potential in Fig. 5. There are about $10^{10}$ myocytes in the heart [a myocyte is about 100 $\mu$m long and 20 $\mu$m diameter, or $\pi.10^{-5}$ mm$^3$ volume and the human adult myocardium occupies a volume of about $4.10^5$ mm$^3$]. Each cell has a slightly different resting potential (e.g. due to small differences in the expression of Na/K pumps) and these transmembrane potential differences become larger during activation, as illustrated for a one-dimensional row of cells in Fig. 5*A*. Since the cells are electrically coupled, the differences between adjacent

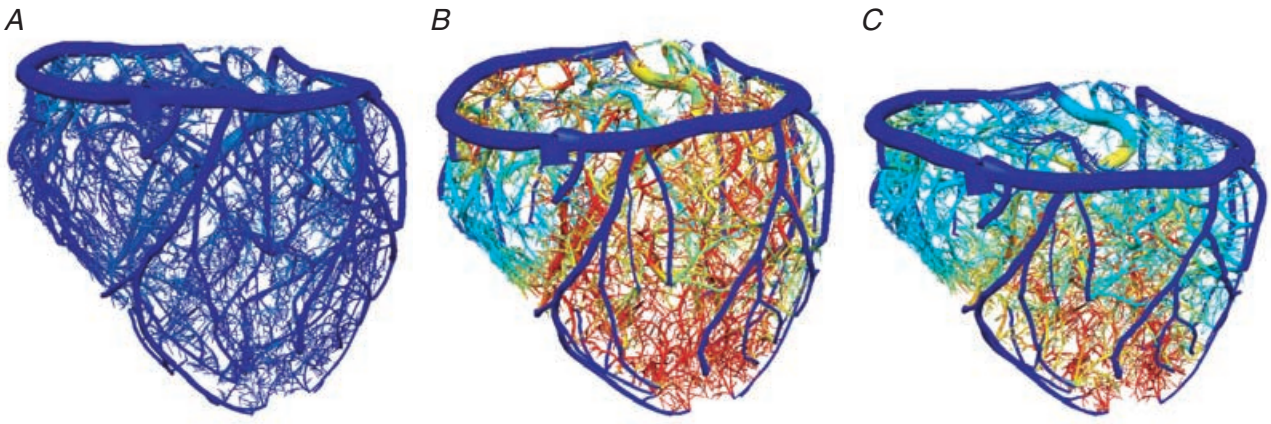


**Figure 11. Model of the coronaries**
The intramyocardial pressure exerted on to the embedded coronary vessels by myocardial contraction. Scaled between 0 KPa (blue) and 8 KPa (red) through the cardiac cycle.

cells are small and to a good approximation the potential field varies smoothly and continuously as shown in Fig. 5*B*. The spatial variation of transmembrane potential is called a *field* and is best represented mathematically by dividing the spatial domain into subdomains called *finite elements*, as shown in Fig. 5*B*. The potential field within an element is then modelled by interpolating values of the potential defined at the *nodes* using the *basis functions* shown in Fig. 5*C*. Since a node is shared between two elements the potential is continuous across element boundaries. If the spatial derivative of potential is also defined at the nodes and the set of basis functions is extended to include interpolation of the derivative (called 'cubic Hermite' basis functions), the finite element representation of the potential field can maintain continuity of its first derivative (and hence current) across the whole domain.

Approximating the discrete cell-by-cell domain with a continuous field representation is the basis for continuum modelling.

Any number of such fields can coexist in the same physical domain. For example, the extracellular potential provides another dependent variable field governed by conservation laws and linked to the transmembrane potential in *bidomain* models of cardiac electrophysiology.

## Model of ventricular geometry

The one-dimensional cubic Hermite basis functions illustrated in Fig. 5 can be extended to three dimensions and used to fit a finite element model of ventricular geometry, as shown in Fig. 6. Anatomically based models have now been developed for the dog (Nielsen *et al.* 1991), pig (Stevens & Hunter, 2003) and rabbit heart (Vetter & McCulloch, 2000).

## Cardiac structure

In addition to the geometry, fundamental to predicting whole organ function, is an accurate representation of the spatial variation in tissue properties which are often based on the underlying microstructure. This fibrous-sheet structure of the heart is illustrated in Fig. 7. The 'fibre' axis is aligned with the local myocytes (i.e. the direction of sarcomere length changes), the sheet axis is orthogonal to the fibre axis, and is in the plane in which myocytes are tightly bound by endomysial collagen into layers 3–4 cells thick. The third 'sheet-normal' axis is orthogonal to this plane. The fibrous-sheet structure for the whole myocardium is illustrated in Fig. 8

## Adding function to the continuum framework

To predict whole organ function such as myocardial mechanics or activation requires the application of physical conservation laws which couple the discrete cellular models into a common spatial framework. These laws are expressed mathematically in the form of integral equations or PDEs (partial differential equations) which balance forces (expressed per unit area as stress) or fluxes (such as an electrical current). These equations are derived from the basic physical laws of nature

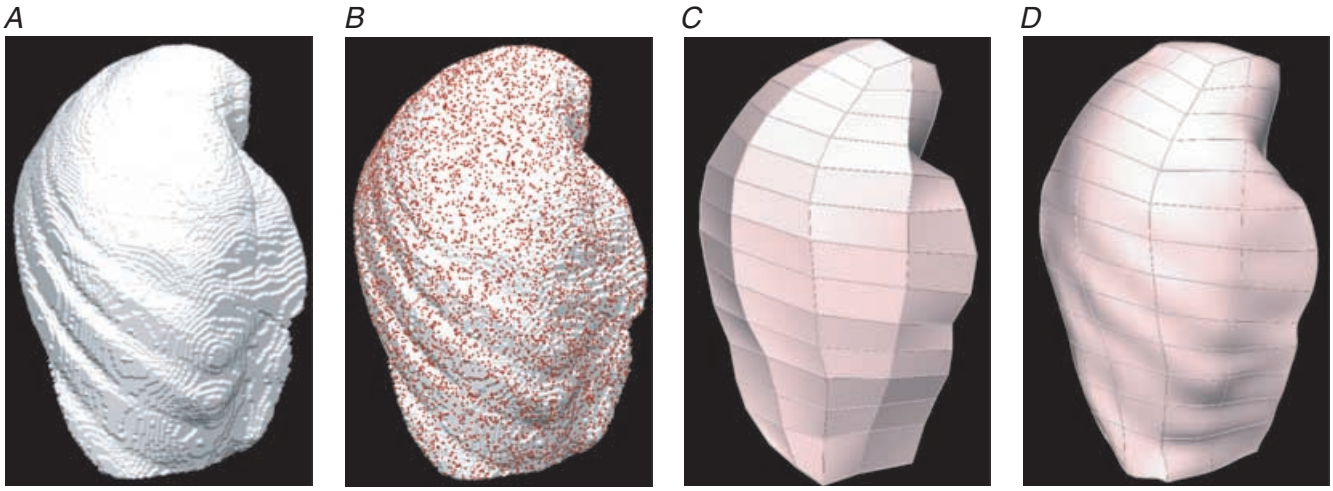


**Figure 12. Extraction of data from CT images, and fitting of a high order finite element mesh for a human lung lobe**
*A,* isosurfaces drawn at the lobe boundary. *B,* random data points calculated to sit on the iso-surfaces. *C,* initial linear volume mesh. *D,* high order mesh fit to the extracted data.

and are independent of the particular properties which characterize the material under consideration (such as soft biological tissue). To solve these balance equations for a particular material requires another equation, called a material law or constitutive equation, which is derived from experimental observation of the particular material. For example, in the case of simulating myocardium contraction, to solve the stress equilibrium equations requires an additional material law linking components of stress to the components of material strain (deformation gradients). The experimentally obtained parameters in this relation are the elastic constants of the tissue. For conservation of current the material law links current to voltage gradients with parameters that express the tissue conductivity.

The systems of equations governing each function are typically non-linear and are solved using numerical methods which reflect the spatial-temporal scale of the underlying phenomena. The rapid upstroke in the action potential of the cardiac myocyte produces large spatial gradients which require representation at high spatial resolution (see Fig. 9). The relative efficiency of the finite difference technique or low order finite element method lends itself to this application. Conversely the continuous stress and strain fields generated by cardiac deformation are effectively predicted using finite elements with high order C1 continuous basis functions. The coupling of systems with large differences in the solution scales provides much of the challenge for modelling computational modelling at this scale in the context of the Physiome Project. Two examples of system coupled modelling are given below.

### Coupled excitation-contraction coupling in myocardial tissue

The electrophysiology model of Noble *et al.* (1998), linked via calcium kinetics to thin filaments kinetics and active force generation by Nickerson *et al.* (2001), provides the voltage source and tension dynamics at a cellular level. Stretch activated channels in the electrophysiological model produce a reverse coupling where mechanical stretch alters the cell action potential. Smith *et al.* (2003*a*) have embedded this cellular model in a grid of finite difference points in a two dimensional finite element model of the cardiac ventricles (see Fig. 10). Excitation was initiated by applying a spatial pattern of stimulus currents based on measurements of Durrer *et al.* (1970), inducing

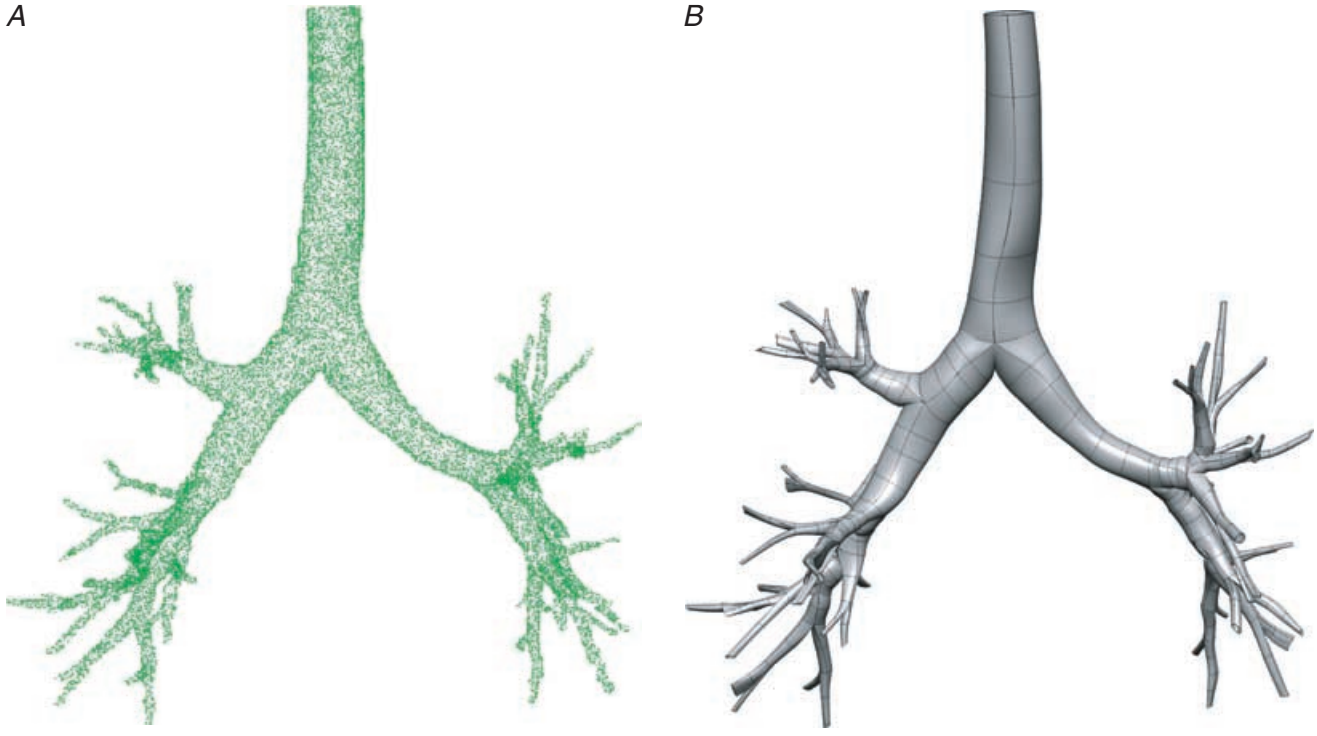


**Figure 13. Fitting surface meshes to human central airways**
*A,* data extracted from CT images. *B,* high order surface mesh fit to the data.

the spread of an activation wave. From this excitation solution at each time step the active tension produced by cell excitation was calculated at the gauss points in the finite element model. Using the non-linear constitutive law, and applying the governing equations, the finite element method was used to calculate tissue deformation in this coupled model. For well characterized networks, in which the relevant genes have been identified and the wiring diagram determined, biophysically realistic non-linear kinetic functions can be assumed for the regulatory interactions. This allows quantitative prediction of transcription rates, etc. in response to perturbations of the network. Kinetic parameters can be determined for individual reactions (for example the binding of transcription factors, degradation rates of mRNAs, etc.) using time-course data from microarrays, or GFP reporter gene approaches (Ronen *et al.* 2002). Interactions between cells are critical in determining patterns of differentiation into distinct tissue types. Models of spatially extended systems must therefore incorporate mechanisms for signalling between cells, transport of gene products as signalling molecules and signalling networks within cells (Mjolsness *et al.* 1991; von Dassow *et al.* 2000; Davidson *et al.* 2002).

## Coupled myocardial mechanics and coronary blood flow

The anatomically-based model of the largest six generations of the coronary arterial tree generated by Smith *et al.* (2000) has been embedded at material points in the model of the cardiac ventricles presented above. By assuming the functional form of the axial velocity profile,

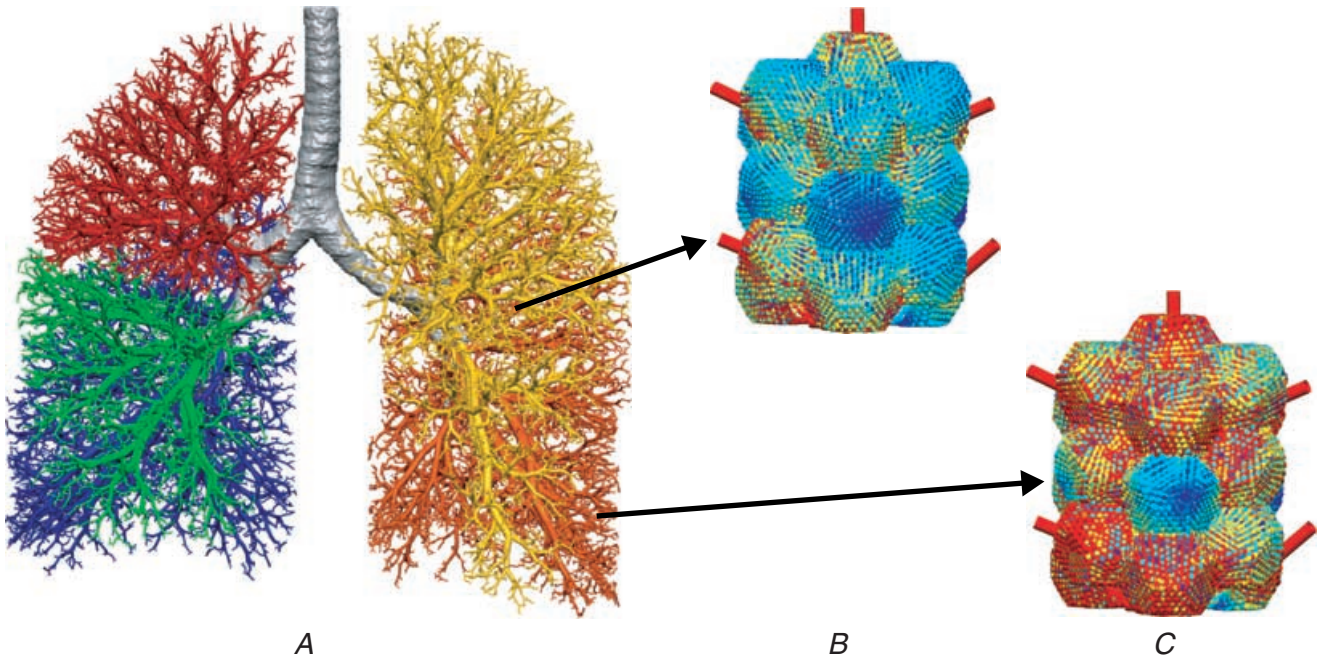


**Figure 14. Airway models**
*A,* airway model using CT data to fit airways down to branch generations 6–9, and airway generation algorithm to fill a CT-based volume mesh from the CT airways out to the terminal bronchioles. The right upper lobe airways are green, right middle lobe are red, right lower lobe are blue, left upper lobe are yellow, and left lower lobe are orange. *B* and *C,* a single alveolar sac comprising 19 alveoli clustered around a central duct, with a dense segmented capillary mesh wrapped over the alveoli. Only a single layer of capillaries passes between each pair of adjacent alveoli. The alveolar-capillary mesh has been used by Burrowes *et al.* (2003) to simulate blood cell transit through the pulmonary microcirculation under different pressure conditions typical of the vertical human lung. By altering the pleural, arteriolar, and venule pressures, transit in the different lung 'zones' can be simulated. *B* shows a flow solution for 'zone 3', where the arterial and venous pressure are greater than the alveolar pressure. *C* shows a flow solution in the same geometry for 'zone 2', where the alveolar pressure is less than arterial but greater than venous pressure. In these solutions red shows the greatest flow, and blue the least.

the three dimensional Navier-Stokes equations governing blood flow have been reduced to one dimension to create an efficient model of coronary haemodynamics (Smith *et al.* 2002*b*). Using the myocardial mechanics framework of Nash & Hunter, (2000) the temporal compression and deformation of each vessel segment was determined. The effect of this spatial–temporal variation in intramyocardial stress was then coupled to blood flow via an elastic pressure radius relationship based on local arterial mechanics. This coupled model of myocardial mechanics and coronary blood flow has been used to quantify the effect of myocardial contraction on coronary blood flow (also known as systolic flow impediment) in the heart (Smith *et al.* 2002*b*), as illustrated in Fig. 11.

## Multi-scale modelling in the lungs

The lung performs gas exchange and metabolic functions through the coupling of several subsystems: the conducting airways that transport air to and from the atmosphere and are a major site of defence; the respiratory airways that are covered with alveoli and hence are the site of gas exchange; the pulmonary capillaries that cover the alveolar walls, bringing blood into close contact with alveolar gas; and the pulmonary vasculature that circulates de-oxygenated blood through the capillary bed and returns oxygenated blood to the heart. The larger transport systems can be considered to be 'embedded' within the pulmonary parenchyma, and the entire system to be mechanically coupled through the vast network of fibres that run along the airways and extend into the lung from its surfaces (Weibel, 1984). Each of these subsystems is therefore intimately coupled to the others, both structurally and functionally. Traditional measures of lung function use global indices such as blood gases or volumes measured by spirometry to non-invasively predict the health of the lung, but the normal lung exhibits structural asymmetry and regional variation in ventilation and perfusion that influences these global measures, and therefore significant loss of lung function can occur through disease before it is detected using global measures. Computational modelling in combination with state-of-the-art medical imaging provides a framework for predicting pulmonary function based on individual structure, such that global measures can be interpreted in terms of regional function.

CT (Computed Tomography) imaging has emerged as the modality of choice for imaging the lung, with state-of-the-art CT of high enough resolution to segment the lung fissures and the major airways and blood vessels (Zhang & Reinhardt, 1999; Kiraly *et al.* 2002). Masked CT images can be used directly to fit finite element models of the lobes and segmented airways and vessels. This process is illustrated for the right upper lobe shown in Fig. 12, using masked images from the Lung Atlas project (Li *et al.* 2003). In Fig. 12*A* iso-surfaces have been constructed to trace the lobe surface. A random spread of data points are generated over the iso-surfaces in Fig. 12*B*. In Fig. 12*C* the data points are projected orthogonally to an initial linear volume mesh surface. Figure 12*D* shows the final high-order finite element mesh that is fit to the CT data using CMISS geometric fitting [the finite element program developed by the Auckland Bioengineering group; www.cmiss.org] (Fernandez *et al.* 2003).

The level of structural detail required in a model of the airways will depend upon its area of application. For example, the details of 3D structure are important for simulating flow fields and particle deposition (Nowak *et al.* 2003), whereas a 1D model that integrates cross-sectional information is considered appropriate for simulating mixing of inspired inert gases (Paiva & Engel, 1979). A further example is in the study of airway collapse, where describing the surface geometry is important but detailed structure can be neglected. An individual CT data set can be used to generate a set of airway models that are patient-specific and relate directly to one another. Volume (3D), surface (2D), or line (1D) finite element airway models can be derived from the single iso-surface data set shown in Fig. 13*A*:

**1** The surface points can be triangulated, and the volume filled with a very fine tetrahedral mesh. This is the traditional approach used for flow modelling, where the surface geometry detail is considered important. However, this produces a very large mesh and consequently flow simulations are generally carried out for only a few airway generations.

**2** Surface meshes can be constructed either by simply triangulating the surface data points (as in 1), or by using the geometry fitting process described for the lobes to fit a high-order surface mesh (Fig. 13*B*).

**3** The surface mesh can be reduced to a 1D mesh by creating a node at each airway bifurcation or ending, and calculating derivatives that force the 1D elements to pass through or very close to the centreline locations of the 2D mesh. Area information can be incorporated by calculating the cross-sectional area at discrete locations down each airway, then using this 'data' to fit area values and derivatives.

These multidimensional models can be used in combination for a single problem (e.g. combining 3D and 1D airflow in different airway generations) or coupled for

solving multiple problems simultaneously (e.g. coupled 3D airflow and 2D wall mechanics).

Current segmentation algorithms can uniquely identify individual human airways from CT imaging down to the 6th to 9th branch generation (Kiraly *et al.* 2002). Beyond this level it becomes increasingly difficult to trace the path of the individual airways. Symmetric (Weibel, 1963) or Horsfield-based (Horsfield *et al.* 1971) models could be coupled to the CT-based airway meshes, but such an approach would not relate the model airways to a spatial location, thus limiting the utility of the model. A mathematical algorithm (Tawhai *et al.* 2000) can be used to generate airways from the fit airway mesh to the terminal bronchioles. The 3D bifurcation-distributive algorithm generates model airways into accurate lobar volume meshes so that the airway geometry and branching pattern depends directly on the lobe geometry. Airways generated into an anatomically -based lobe geometry are shown in Fig. 14*A*, with the airways coloured to differentiate between the five lobes.

The alveoli in the respiratory airways form a mechanically-tethered elastic tissue. The alveoli are densely packed, with multifaceted shapes (Weibel, 1984). The respiratory tissue can be modelled as a 3D structure for mechanics studies (Denny & Schroter, 1996; Burrowes *et al.* 2003), or as a 1D structure for simulating inert gas mixing (Verbanck & Paiva, 1990; Dutrieue *et al.* 2000). For example, Burrowes *et al*. (2003) have modelled the alveolar geometry using a space-filling Voronoi mesh. This approach fills a prescribed volume with space-filling alveoli and central duct spaces, and has been used further by Burrowes *et al*. (2003) as a base structure over which a model of the pulmonary microcirculation was generated (Fig. 14*B*). The 3D mesh has an anatomically consistent geometry and forms a structural framework for incorporating the distribution of extracellular collagen, elastin, and basement membrane for mechanical analysis.

By utilizing a hierarchy of models for the pulmonary system, whether structural (e.g. 3D, 2D, and 1D airway models) or functional (e.g. using a multibranching model to parameterize a lumped model for inert gas mixing in the acinus) we have a powerful modelling framework that allows us to move between, and link, different levels of interest for simulating a range of functional problems (Tawhai & Hunter, 2001*a*,*b*). An example of applying this approach to link CT-imaged regional information to global measures is by investigating the sensitivity of predicted global measures to changes in model geometry: do the models predict that global measures will be influenced by geometry changes within the 'normal' range, or alternatively, do model predictions suggest that geometry changes outside the 'normal' range would not be detected by the global measures? This investigation requires incorporation of ventilation distribution, perfusion distribution, soft tissue mechanics, inert gas mixing to simulate transport of the imaged contrast agent, and gas exchange. Each of these components can be modelled in detail (3D air flow, parenchymal micromechanics) or reduced and combined for a tractable model (1D air flow calibrated against 3D, compressible finite elasticity for the air-tissue 'composite'). The effect of respiratory tissue geometry changes can be probed using multibranching models of the acinus, or the effect of conducting airway geometry changes can be probed using anatomically based conducting airway models coupled to simple lumped parameter models of the acinus.

## Conclusions

The canvas on which we have constructed this review is vast. It ranges in scale from modelling molecular processes to modelling whole organs and systems, and its theoretical range encompasses linking physiology to mainstream biological theory in relation to evolution and development. It is necessary to emphasize the enormity of the challenge that quantitative integrative physiology faces for several reasons

First, the magnitude of the challenge is not yet fully appreciated. Integrative physiology has been so systematically overshadowed over recent decades by the developments in molecular biology that its scale and funding have not kept pace. We believe that this balance should, and probably will, progressively change as the power of quantitative integrative work becomes clearer. To the extent that we can show how higher-level functional understanding clarifies molecular understanding, and is even necessary for that understanding, the more this area of science will attract some of the best minds to tackle its challenge.

Second, we wish to emphasize the range of skills that will be required. The authors of this review include a mathematician, a physiologist, a computer scientist and four bioengineers. The project we are describing is necessarily multidisciplinary. It will transform the way in which physiology is done, just as much as the molecular revolution did. Integrative physiology in the past was essentially qualitative. It will in future become a highly quantitative, computer-intensive discipline.

Third, we really are only at the very beginning. There is still a long way to go in the development of the relevant models. In many areas of physiology other

than the heart and lungs, modelling is not yet as well developed. As models improve in sophistication, more gene expression and proteomic data can usefully be incorporated for predictive modelling. As we have noted, spatial information is crucial in making the step from cellular to tissue level properties. Adequately dealing with spatial variation in parameters and properties remains one of the major gaps in current modelling approaches at all spatial levels. Proteomic information on biomolecular interactions will be important for unravelling metabolic and signalling pathways operating in the cell, and in particular in response to disease and injury. Identifying functional interactions in signalling and genetic networks is of particular interest as they regulate the coordinate expression of functional groups of genes, with a few key pathways switching between alternative cell fates (Neves & Iyengar, 2002; Bhalla, 2003). Gene network modelling is also at an early stage in its development. A great deal of attention is currently focused in reverse engineering dynamic models of regulatory networks directly from gene expression and proteomic data (Brazhnik *et al.* 2002; Bolouri & Davidson, 2002; Kholodenko *et al.* 2002; Tegner *et al.* 2003). This may be facilitated by new results which indicate that in the future it may be possible to manipulate the genome *in situ*, and by selectively turn genes on and off *in vivo* using genetically engineered switches and other control mechanisms (Gardner *et al.* 2000; Elowitz & Leibler, 2000; Hasty *et al.* 2002). If reverse engineering can be successfully accomplished and automated then the analysis of highthroughput data will feed directly into parameterization of dynamic models which can be used for quantitative prediction of the regulatory networks which underly physiological function and the response of tissue to injury and disease.

One of the intriguing opportunities presented by the availability of high resolution imaging and anatomically based computational models is that of the patient-specific modelling. That is, the generic model of the heart or lungs, shown in Fig. 6 and Fig. 14, respectively, can be adjusted to match MR images of the heart or helical scan CT images of the lungs. Coupled with measurements of both gene sequence and physiological function for that individual, the realization of patient-specific, model-based clinical diagnosis becomes more feasible.

In this article we have surveyed issues of complexity and modelling from genes and proteins to tissues and organs. A framework for the physiome project is being put in place which involves developing ontologies for describing the biological knowledgebase, markup languages for encapsulating models of structure and function at all spatial scales. The greatest challenges at present are developing the means of linking models across the spatial scales.

What are the factors that will continue to drive this project? One of these is the sheer excitement of unravelling the logic of life at all levels of complexity. This intellectual challenge will become even greater as the various pieces are put together. The more of the jigsaw puzzle we see, the more we will want to complete it. A second important factor will be clinical and pharmaceutical relevance. The development of the 'heart physiome' as the first physiome model of an organ was motivated by the obvious clinical importance of heart disease.

## Appendix

### Ontologies and Tools

The IUPS physiome project is developing a number of tools for biological modelling. As we have discussed, ontologies are fundamental to these. How a particular tool makes use of an ontology depends on what information it needs from it. This process can be viewed in a number of ways:

**1** defining machine interpretable representation languages

**2** building internal data structures for databases or object orientated constructs

**3** representing datasets for simulation and visualization

**4** finding results of complex queries over ontological domains

**5** generating data for dynamic navigation interfaces that people use to browse, select, and manage representations

This is certainly not an exhaustive list, but does give us the right perspective to interpret the main ontological domains we are building or integrating (Fig. 2). A good way to get a feeling for the value of building and integrating these ontology domains is to consider some use-cases from the perspective of a biologist or modeller using various tools. We can envisage the goal of a biologist as locating models, running simulations, and visualizing systems, and perhaps identifying systems that fit their data sets, and the goal of a modeller to be investigating the theoretical foundations of models, building models, comparing models, and like the biologist, running simulations and visualizing the systems.

### Examples of how the ontologies, markup languages and tools are used by physiologists and modellers are illustrated here

**Use-case 1.** A biologist, using an interface to the BioPAX ontology, locates the cAMP/PKA signalling cascade that participates in the regulation of L-type calcium channel

activity. From this concept they locate CellML models that describe this system, and are able to run simulations, manipulate these, and visualize their behaviour.

**Use-case 2.** A biologist locates L-type calcium channels through an anatomical navigation interface. From here they can investigate the 3D structure of the channel, its physiological function, or publications that relate to it. Each of these steps helps to gather or filter a set of models with which they can continue with simulation and visualization.

**Use-case 3.** A biologist with a protein domain motif, perhaps with identifiers from the Gene Ontology or protein interaction databases, obtains a set of models that refer to this motif. From here they can look at 3D structures, physiological function, or visualizations of its behaviour in various models.

**Use-case 4.** A biologist with a data set wants to find a model that could help them interpret data from experiments. Using navigation or query interfaces, they can find a set of models that contain the correct entities, or describe the appropriate physiological process, or use particular modelling theories. From here they enter an iterative process of model fitting and system identification, i.e. reducing the set of models to those that provide useful levels of accuracy. The parameter data sets and the raw data sets themselves can be submitted for peer review to be included in the repositories for other people to use.

**Use-case 5.** A modeller has located a particular model. They are able to run it in a simulation, visualize its behaviour, interpret the mathematical theories it was built from, and then edit it in a model editor. They are able to submit annotations to the original model, or submit new models for peer review to be included in the repositories and ontologies.

**Use-case 6.** A modeller starts with a publication, obtains a set of models that describe both the publication and unpublished models of the same processes. They can view comparisons of these models that highlight the similarities and differences in architecture and modelling theories used by the modellers who created them.

**Use-case 7.** A modeller has a particular goal in mind, in this case, coupling their model to models that describe systems at a finer physiological scale to theirs. They can find a concept of coupling scales in the navigation interface that interfaces with the modelling theory ontology. From here they see mathematical systems or examples for coupling between scales, and through these select actual models that implement these. They now select subsystems from the library, or make up their own, and have the option of selecting model templates that help them to couple the subsystems into their model. For example, selecting various subsystems that describe the signalling pathway leading to the activation of L-type calcium channels, and integrating these into their continuum model that may couple a spatial variation of activation of $\beta 1$ and $\beta 2$

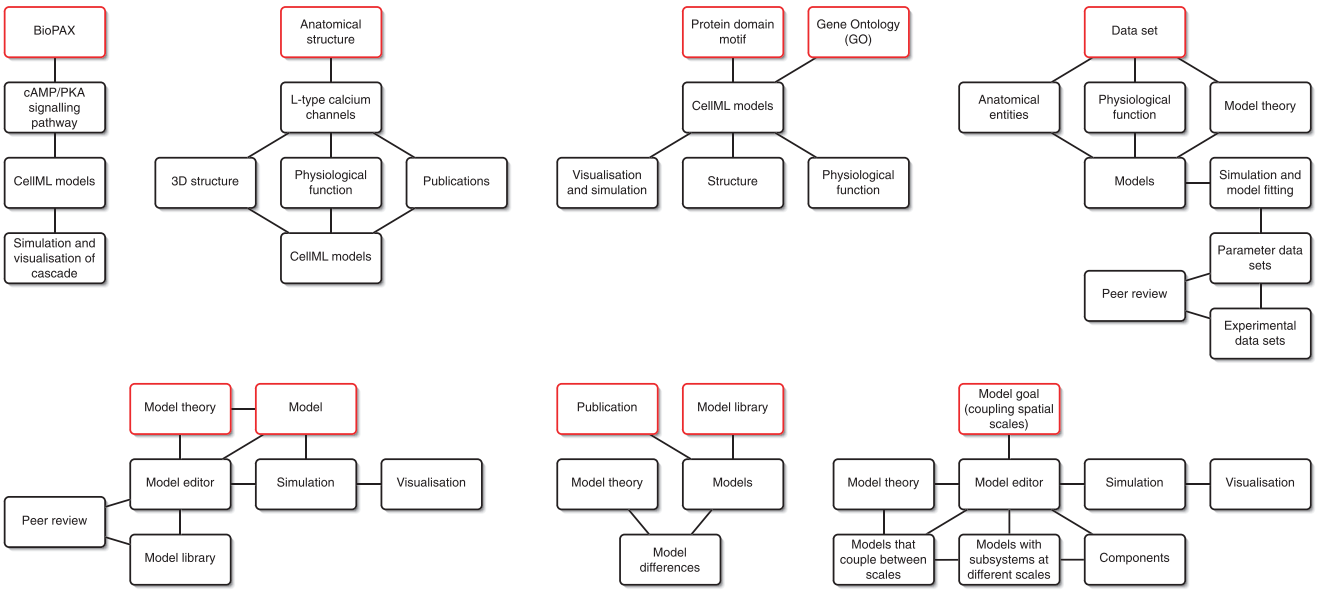


**Figure A1. Use-cases to illustrate ways in which models are accessed**

adrenergic receptors and the resulting spatio-temporal propagation of activation of the muscle.

## References


Alon U, Surette MG, Barkai N & Leibler S. (1999). Robustness in bacterial chemotaxis. *Nature* **397**, 168–171.

Alter O, Brown PO & Botstein D (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci* **97**, 10101–10106.

Antzelevitch C, Nesterenko VV, Muzikant AL, Rice JJ, Chien G & Colatsky T (2001). Influence of transmural gradients on the electrophysiology and pharmacology of ventricular myocardium. Cellular basis for the Brugada and long–QT syndromes. *Philosoph Transactions Royal Soc, Series A* **359**, 1201–1216.

Barrans JD, Allen PD, Stamatiou D, Dzau VJ & Liew CC (2002). Global gene expression profiling of endstage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* **160**, 2035–2043.

Barrans JD, Stamatiou D & Liew CC (2001). Construction of a human cardiovascular cDNA microarray: Portrait of the failing heart. *Biochem Biophys Res Comms* **280**, 964–969.

Bhalla US (2003). Understanding complex signaling networks through models and metaphors. *Prog Biophys Mol Biol* **81**, 45–65.

Bober M, Wiehe K, Yung C, Suzek TO, Lin M, Baumgartner W Jr. & Winslow RL (2002). Cage: Cardiac gene expression knowledgebase. *Bioinformatics* **18**, 1013–1014.

Bolouri H & Davidson EH (2002). Modeling transcriptional regulatory networks. *Bioessays* **24**, 1118–1129.

Brazhnik P, De La Fuente A & Mendes P (2002). Gene networks: how to put the function in genomics. *Trends Biotechnol* **20**, 467–472.

Brenner S (1998). Biological Computation. In: *The Limits of Reductionism in Biology*.Vol. 213, ed. Novartis Foundation, pp. 106–116. Wiley, London.

Burrowes KS, Tawhai MH & Hunter PJ (2003). Modeling RBC and neutrophil distribution through an anatomically based pulmonary capillary network. *Ann Biomed Engineering* (in press).

Churchill GA (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genet* **32** (Suppl), 490–495.

Clancy CE & Rudy Y (2002). $Na^+$ channel mutation that causes both Brugada and long–QT syndrome phenotypes: a simulation study of mechanism. *Circulation* **12**, 1208–1213.

Claverie JM (2001). Gene number – what if there are only 30,000 human genes? *Science* **291**, 1255–1257.

Cohn JN, Ferrari R & Sharpe N (2000). Cardiac remodeling – concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. *J Am College Cardiol* **35**, 569–582.

Collins FS, Green ED, Guttmacher AE & Guyer MS (2003). A vision for the future of genomics research. *Nature* **422**, 835–847.

Cook SA & Rosenzweig A (2002). DNA microarrays: Implications for cardiovascular medicine. *Circ Res* **91**, 559–564.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, Clarke PJC, Rust AG, Pan ZJ, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L & Bolouri H (2002). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev Biol* **246**, 2–13.

Dawkins R (1976). *The Selfish Gene*. Oxford University Press, Oxford, UK.

Dennis C (2002). The brave new world of RNA. *Nature* **418**, 122–124.

Denny E & Schroter RC (1996). A mathematical model for the morphology of the pulmonary acinus. *J Biomechan Engineering* **118**, 210–215.

Diamond JM (1993). Evolutionary Physiology. In *The Logic of Life*, ed Boyd CAR & Noble D, pp 89–111. Oxford University Press, Oxford, UK.

DiFrancesco D & Noble D (1985). A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Phil Trans R Soc Lond B Biol Sci* **10**, 353–398.

Durrer D, Van Dam RT, Freud GE, Janse MJ, Meijler FL & Arzbaecher RC (1970). Total excitation of the isolated human heart. *Circulation* **41**, 899–912.

Dutrieue B, Van Holsbeeck F, Verbanck S & Paiva M (2000). A human acinar structure for simulation of realistic alveolar plateau slopes. *J Appl Physiol* **89**, 1859–1867.

Eisen MB, Spellman PT, Brown PO & Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* **95**, 14863–14868.

Elowitz MB & Leibler S (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang ZP, Goldstein SR, Weiss RA & Liotta LA (1996). Laser capture microdissection. *Science* **274**, 998–1001.

Fernandez JW, Mithraratne P, Tawhai MH & Hunter PJ (2003). Geometric fitting and customization of anatomically based cubic Hermite finite element volume meshes. *Biomech Model Mechanobiol* (in press).

Gardner TS, Cantor CR & Collins JJ (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.

Gould SJ (2002). *The Structure of Evolutionary Theory*. Harvard, Cambridge, MA, USA.

Guccione JM & McCulloch AD (1993). Mechanics of active contraction in cardiac muscle: Part I-Constitutive relations for fiber stress that describe deactivation. *J Biomech Eng* **115**(1), 72–81.

Hasty J, McMillan D & Collins JJ (2002). Engineered gene circuits. *Nature* **420**, 224–230.

Henriksen PA & Kotelevtsev Y (2002). Application of gene expression profiling to cardiovascular disease. *Cardiovasc Res* **54**, 16–24.

Hodgkin AL & Huxley AF (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117**, 500–544.

Holter NS, Mitra M, Maritan A, Cieplak M, Banavar JR & Fedoroff NV (2000). Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc Natl Acad Sci* **97**, 8409–8414.

Hooks DA, Tomlinson KA, Marsden SG, Legrice IJ, Smaill BH, Pullan AJ & Hunter PJ (2002). Cardiac microstructure: Implications for electrical propagation and defibrillation in the heart. *Circ Res* **9**, 331–338.

Horsfield K, Dart G, Olsen DE, Filley GF & Cumming G (1971). Models of the human bronchial tree. *J Appl Physiol* **31**, 207–217.

Hunter PJ & Borg TK (2003). Integration from proteins to organs: The Physiome Project. *Nature Reviews Molecular and Cell Biology* **4**, 237–243.

Hunter PJ, Kohl P & Noble D (2001). Integrative models of the heart: achievements and limitations. *Phil Trans R Soc Lond A* **359**, 1049–1054.

Hunter PJ, McCulloch AD & Ter Keurs HEDJ (1998). Modelling the mechanical properties of cardiac muscle. *Prog Biophys Molec Biol* **69**, 289–331.

Hunter PJ, Robbins P & Noble D (2002). The IUPS Human Physiome Project. *European J Physiol* **445**, 1–9.

International Human Genome Mapping Consortium (2001). A physical map of the human genome. *Nature* **409**, 934–941.

Jafri MS, Rice JJ & Winslow RL (1998). Cardiac $Ca^{2+}$ dynamics: the roles of ryanodine receptor adaptation and sarcoplasmic reticulum load. *Biophys J* **74**, 1149–1168.

Jeong H, Tombor N, Albert R, Oltval ZN & Barabasi AL (2000). The large-scale organization of metabolic networks. *Nature* **407**, 651–654.

Keller L (1999). *Levels of Selection in Evolution*. Princeton University Press, Princeton, NJ, USA.

Kepler TB & Elston TC (2001). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* **81**, 3116–3136.

Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV & Hoek JB (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci* **99**, 12841–12846.

Kiraly A, Higgins WE, McLennan G, Hoffman EA & Reinhardt JM (2002). Three-dimensional human airway segmentation for clinical virtual bronchoscopy. *Acad Radiol* **9**, 1153–1168.

Krakauer D (2002) From Physics to Phenomenology Levels of description and levels of selection. In *'in Silico' Simulation of Biological Processes*, ed. Novartis Foundation, pp 42–52. Wiley, London.

Li B, Christensen GE, Hoffman EA, McLennan G & Reinhardt JM (2003). Establishing a normative atlas of the human lung: intersubject warping and registration of volume tric CT images. *Acad Radiol* **10**, 255–265.

Liang S, Fuhrman S & Somogyi R (1998). REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium on Biocomputing* **3**, 18–29.

Luo CH & Rudy Y (1991). A model of the ventricular cardiac action potential Depolarization, repolarization, and their interaction. *Circ Res* **68**, 1501–1526.

Luo CH & Rudy Y (1994). A dynamic model of the cardiac ventricular action potential I Simulations of ionic currents and concentration changes. *Circ Res* **74**, 1071–1096.

MacBeath G (2002). Protein microarrays and proteomics. *Nat Genet* **32** (Suppl), 526–532.

Mattick JS (2001). Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Reports* **2**, 986–991.

Mazhari R, Greenstein JL, Winslow RL, Marban E & Nuss HB (2001). Molecular interactions between two long–QT syndrome gene products, HERG and KCNE2, rationalized by in vitro and in silico analysis. *Circ Res* **89**, 33–38.

McAdams HH & Arkin A (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* **15**, 65–69.

Michailova A & McCulloch AD (2001). Model study of ATP and ADP buffering, transport of Ca (2+) and Mg (2+), and regulation of ion pumps in ventricular myocyte. *Biophys J* **81**, 614–629.

Mills JC, Roth KA, Cagan RL & Gordon JI (2001). DNA microarrays and beyond: completing the journey from tissue to cell. *Nature Cell Biol* **3**, E175–E178.

Mjolsness E, Sharp DH & Reinitz J (1991). A connectionist model of development. *J Theoret Biol* **152**, 429–453.

Napoli C, Lerman LO, Sica V, Lerman A, Tajana G & De Nigris F (2003). Microarray analysis: a novel research tool for cardiovascular scientists and physicians. *Heart* **89**, 597–604.

Nash MP & Hunter PJ (2000). Computational mechanics of the heart: from tissue structure to ventricular function. *J Elasticity* **61**, 113–141.

Neves SR & Iyengar R (2002). Modeling of signaling networks. *Bioessays* **24**, 1110–1117.

Nickerson D, Smith NP & Hunter PJ (2001). A model of cardiac cellular electromechanics. *Phil Trans Royal Soc* **A359**, 1159–1172.

Nielsen PMF, Legrice IJ, Smaill BH & Hunter PJ (1991). Mathematical model of geometry and fibrous structure of the heart. *Am J Physiol* **260**, H1365–H1378.

Noble D (2002*a*). Biological Computation. In *Encyclopedia of Life Sciences*. Nature Publishing Group, London.

Noble D (2002*b*). Is the Genome the Book of Life? *Physiol News* **46**, 18–20.

Noble D (2002*c*). Modelling the heart: from genes to cells to the whole organ. *Science* **295**, 1678–1682.

Noble D (2002*d*). Modelling the heart: insights, failures and progress. *Bioessays* **24**, 1155–1163.

Noble D (2002*e*). Unravelling the genetics and mechanisms of cardiac arrhythmia. *Proc Natl Acad Sci, USA* **99**, 5755–5756.

Noble D & Boyd CAR (1993). The Challenge of Integrative Physiology. In *The Logic of Life*, ed Boyd CAR & Noble D, pp 1–13. Oxford University Press, Oxford.

Noble D & Rudy Y (2001). Models of cardiac ventricular action potentials: iterative interaction between experiment and simulation. *Phil Trans R Soc Lond A* **359**, 1127–1142.

Noble D, Varghese A, Kohl P & Noble P (1998). Improved guinea-pig ventricular cell model incorporating a diadic space, IKr and IKs, and length- and tension-dependent processes. *Can J Cardiol* **14**, 123–134.

Nowak N, Kakade PP & Annapragada AV (2003). Computational fluid dynamics simulation of airflow and aerosol deposition in human lungs. *Ann Biomed Engineering* **31**, 374–390.

Paiva M & Engel LA (1979). Pulmonary interdependence of gas transport. *J Appl Physiol* **47**, 296–305.

Panerai RB (1980). A model of cardiac muscle mechanics and energetics. *J Biomech* **13**, 929–940.

Patino WD, Mian OY, Shizukuda Y & Hwang PM (2003). Current and future applications of SAGE to cardiovascular medicine. *Trends Cardiovascular Med* **13**, 163–168.

Ronen M, Rosenberg R, Shraiman BI & Alon U (2002). Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci* **99**, 10555–10560.

Saucerman JJ, Brunton LL, Michailova AP & McCulloch AD (2003). Modeling beta-adrenergic control of cardiac myocyte contractility in silico. *J Biol Chem* **278**, 47997–48003.

Schnell S & Mendoza C (1997). Theoretical description for polymerase chain reaction. *J Theoret Biol* **188**, 313–318.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, Goddard A, Yang RH, Jin HK & Lowe DG (2000). Application of cdna microarrays in determining molecular phenotype in cardiac growth & development and response to injury. *Circulation* **101**, 1990–1999.

Simkhovich BZ, Kloner RA, Poizat C, Majoram P & Kedes LH (2003). Gene expression profiling – a new approach in the study of myocardial ischemia. *Cardiovasc Pathol* **12**, 180–185.

Slonim DK (2002). From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* **32** (Suppl), 502–508.

Smirnov SV & Aaronson PI (1994). Alteration of the transmembrane K+ gradient during development of delayed rectifier in isolated rat pulmonary arterial cells. *J General Physiol* **104** 241–264.

Smith NP (2003). From sarcomere to cell: an efficient algorithm for linking mathematical models of muscle contraction. *Bull Math Biol* **65**, 1141–1162.

Smith NP, Buist ML & Pullan AJ (2003*a*). Altered T wave dynamics in a contracting cardiac model. *J Cardio Electrophys* **14**, S203–S209.

Smith, NP, Pullan AJ & Hunter PJ (2000). The generation of an anatomically accurate geometric coronary model. *Ann Biomed Engineering* **28**, 14–25.

Smith NP, Mulquiney PJ, Nash MP, Bradley CP, Nickerson DP & Hunter PJ (2002). Mathematical modeling of the heart: cell to organ. *Chaos Solutions Fractals* **13**, 1613–1621.

Smith NP, Pullan AJ & Hunter PJ (2002). An anatomically based model of transient coronary blood flow in the heart. *SIAM J App Math* **62**, 990–1018.

Smith NP, Stevens C & Hunter PJ (2003*b*). Modelling computational ventricular mechanics. *Appl Mechanics Rev* (in press).

Snyders DJ & Chaudhary A (1996). High affinity open channel block by dofetilide of HERG expressed in a human cell line. *Mol Pharmacol* **49**, 949–955.

Speed TP (2003). *Statistical Analysis of Gene Expression Microarray Data*. Chapman & Hall, London.

Stanton LW, Garrard LJ, Damm D, Garrick NL, Lam A, Kapoun AM, Zheng Q, Protter AA, Schreiner GF & White RT (2000). Altered patterns of gene expression in response to myocardial infarction. *Circ Res* **86**, 939–945.

Stevens C & Hunter PJ (2003). Sarcomere length changes in a model of the pig heart. *Prog Biophys Molec Biol* **82**, 229–241.

Stevens C, Remme E, Legrice IJ & Hunter PJ (2003). Ventricular mechanics in diastole: material parameter sensitivity. *J Biomechanics* **36**, 737–748.

Stoeckert CJ Jr., Causton HC & Ball CA (2002). Microarray databases: standards and ontologies. *Nat Genet* **32** (Suppl), 469–473.

Tawhai MH & Hunter PJ (2001*a*). Characterising respiratory airway gas mixing using a lumped parameter model of the pulmonary acinus. *Respiration Physiol* **127**, 241–248.

Tawhai MH & Hunter PJ (2001*b*). Multibreath washout analysis: modelling the influence of conducting airway asymmetry. *Respiration Physiol* **127**, 249–258.

Tawhai MH, Pullan AJ & Hunter PJ (2000). Generation of an anatomically based three-dimensional model of the conducting airways. *Ann Biomed Engineering* **28**, 793–802.

Taylor CF, Paton NW, Garwood KL, Kirby PD, Stead DA, Yin ZK, Deutsch EW, Selway L, Walker J, Riba-Garcia I, Mohammed S, Deery MJ, Howard JA, Dunkley T, Aebersold R, Kell DB, Lilley KS, Roepstorff P, Yates JR, Brass A, Brown AJP, Cash P, Gaskell SJ, Hubbard SJ & Oliver SG (2003). A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nature Biotechnol* **21**, 247–254.

Tegner J, Yeung MKS, Hasty J & Collins JJ (2003). Reverse engineering gene networks: Integrating genetic perturbations with dynamic modeling. *Proc Natl Acad Sci* **100**, 5944–5949.

The Chipping Forecast II (2002). *Nature Genetics* **32**. (This supplement to Nature Genetics contains a number of review and survey articles on microarray technology and applications).

Tomlinson KA, Hunter PJ & Pullan AP (2002). A finite element method for an eikonal equation model of myocardial excitation wavefront propagation. *SIAM J Appl Maths* **63**, 324–350.

Ushimaru M & Fukushima Y (2002). Ca (2+) transport of Ca (2+)-ATPase of the sarcoplasmic reticulum in an ADPsensitive phosphoenzyme state. *J Biochem Mol Biol Biophys* **6**, 101–106.

Velculescu VE, Zhang L, Vogelstein N & Kinzler KW (1995). Serial analysis of gene expression. *Science* **270**, 484–487.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.* ( 2001). The sequence of the human genome. *Science* **291**, 1304–1351.

Verbanck S & Paiva M (1990). Model simulations of gas mixing and ventilation distribution in the human lung. *J Appl Physiol* **69**, 2269–2279.

Vetter FJ & McCulloch AD (2000). Three dimensional stress and strain in passive rabbit left ventricle: a model study. *Ann Bioeng* **28**, 781–792.

von Dassow G, Meir E, Munro EM & Odell GM (2000). The segment polarity network is a robust developmental module. *Nature* **406**, 188–192.

Weibel ER (1963). *Morphometry of the Human Lung*. Spriner-Verlag, Berlin.

Weibel ER (1984). *The Pathway for Oxygen: Structure and Function in the Mammalian Respiratory System*. Harvard, Cambridge, MA, USA.

Williams GC (1992). *Natural Selection: Domains, Levels and Challenges*. Oxford University Press, Oxford.

Winslow RL & Boguski MS (2003). Genome informatics – Current status and future prospects. *Circ Res* **92**, 953–961.

Winslow RL, Rice JJ & Jafri S (1998). Modeling the cellular basis of altered excitation-contraction coupling in heart failure. *Prog Biophysics Mol Biol* **69**, 497–514.

Wong AY (1972). Mechanics of cardiac muscle, based on Huxley's model: simulation of active state and force-velocity relation. *J Biomech* **5**, 107–117.

Yamamoto M, Wakatsuki T, Hada A & Ryo A (2001). Use of serial analysis of gene expression (SAGE) technology. *J Immunol Meth* **250**, 45–66.

Yang J, Clark JW, Bryan RM & Robertson C (2003). The myogenic response in isolated rat cerebrovascular arteries: smooth muscle cell model. *Med Eng Phys* **25**, 691–709.

Yeung MKS, Tegner J & Collins JJ (2002). Reverse engineering gene networks using singular value decomposition and robust regression. *Proc Natl Acad Sci* **99**, 6163–6168.

Zhang L & Reinhardt JM (1999). Detection of lung lobar fissures using fuzzy logic. *Proc SPIE* **3660**, 188–199.

## Acknowledgements

PJH gratefully acknowledges many contributions by colleagues in the Auckland Bioengineering Institute and the support of grants from the NZ Royal Society (to the Center for Molecular Biodiscovery), the NZ Foundation for Research, Science and Technology, and the Wellcome Trust. MHT acknowledges the input of the Lung Atlas collaborators and, in particular, Eric Hoffman, Joe Reinhardt and Juerg Tschirren.