Experimental Physiology – *Review Article*

The Cardiac Physiome: perspectives for the future

James Bassingthwaighte[1], Peter Hunter[2] and Denis Noble[3]

[1]*Bioengineering, N210G Foege Building, 1705 NE Pacific Street, University of Washington, Box 35-5061 Seattle, WA 98195-5061, USA*
[2]*University of Auckland, Auckland Bioengineering Institute (ABI), 70 Symonds Street, Auckland, New Zealand*
[3]*Department of Physiology, Anatomy & Genetics, Parks Road, Oxford, OX1 3PT, UK*

The Physiome Project, exemplified by the Cardiac Physiome, is now 10 years old. In this article, we review past progress and future challenges in developing a quantitative framework for understanding human physiology that incorporates both genetic inheritance and environmental influence. Despite the enormity of the challenge, which is certainly greater than that facing the pioneers of the human genome project 20 years ago, there is reason for optimism that real and accelerating progress is being made.

Corresponding author D. Noble: Department of Physiology, Anatomy & Genetics, Parks Road, Oxford OX1 3PT, UK. Email: denis.noble@dpag.ox.ac.uk

The Physiome Project was formally launched at a satellite symposium of the International Union of Physiological Sciences (IUPS) Congress in St Petersburg in 1997. Just over a decade later, where are we? Are the aims and principles outlined at that time being fulfilled? In this article, we address these questions by discussing multiscale analysis and modularity in biological systems, the various approaches to mathematical analysis in biology, and the framework being established by the IUPS Physiome Project to help with the understanding of complex physiological systems through the use of biophysically based mathematical models that link genes to organisms.

Multiscale analysis

One of the central principles is that complex systems like the heart are inevitably multiscalar, composed of elements of diverse nature, constructed spatially in a hierarchical fashion. This requires linking together different types of modelling at the various levels. It is neither possible nor explanatory to attempt to model at the organ and system levels in the same way as at the molecular and cellular levels. To represent the folding, within microseconds, of a single protein using quantum mechanical calculations requires months of computation on the fastest existing parallel computers (such as IBM's Blue Gene). It would require unbelievably large numbers of such computers (one estimate is 10^{27}; Noble, 2006) to analyse just a single cell in this degree of detail. Even if we could do it, we would still need to abstract from the mountain of computation some explanatory principles of function at the cellular level. Furthermore, we would be completely lost within that mountain of data if we did not include the constraints that the cell as a whole exerts on the behaviour of its molecules. This is the fundamental reason for employing the middle-out approach. In multiscalar systems with feedback and feedforward loops between the scale levels, there may be no privileged level of causation (Noble, 2008*a*).

The impressive developments in epigenetics over the last decade (Bird, 2007) have reinforced this conclusion by revealing the nature and extent of some of the molecular mechanisms by which the higher level constraints are exerted. In addition to regulation by transcription factors, the genome is extensively marked by methylation and binding to histone tails. It is partly through this marking that a heart cell achieves, with precisely the same genome, the distinctive pattern of gene expression that makes it a heart cell rather than, for example, a bone cell or a liver cell. The marking is transmitted down the cell lines as they divide to form more cells of the same kind. The feedbacks between physiological function and gene expression that must be responsible are still to be discovered. Since fine gradations of expression underlie important regional characteristics of cardiac cells, making a pacemaker cell different from a ventricular cell, and making different ventricular cells have different repolarization times, this must be one of the important targets of future work on the Cardiac Physiome. We need to advance beyond annotating those gradients of expression to understanding how they arise during development and how they are

maintained in the adult. This is one of the ways in which quantitative physiological analysis will be connected to theories of development and of evolution. The logic of these interactions in the adult derives from what made them important in the process of natural selection. Such goals of the Physiome Project may lie far in the future, but they will ultimately be important in deriving comprehensive theories of the 'logic of life'.

A second reason why multiscale analysis is essential is that a goal of systems analysis must be to discover at which level each function is integrated. Thus, pacemaker activity is integrated at the cell level; single sinus node cells show all the necessary feedback loops that are involved. Below this level, it does not even make sense to speak of cardiac rhythm. At another level, understanding fibrillation requires analysis at least at the level of large volumes of tissue and even of the whole organ. Likewise, understanding the function of the heart as a mechanical pump is, in the end, an organ-level property. Another way of expressing this point is to say that high-level functions are emergent properties that require integrative analysis and a systems approach. The word 'emergent' is itself problematic. These properties do not 'emerge' blindly from the molecular events; they were originally guided by natural selection and have become hard-wired into the system. Perhaps 'system properties' would be a better description. They exist as a property of the system, not just of its components.

A third reason why multiscale analysis is necessary is that there is no other way to circumvent the 'genetic differential effect problem' (Noble, 2008*b*). This problem arises because most interventions at the level of genes, such as gene knockouts and mutations, do not have phenotypic effects. The system as a whole is very effective in buffering genetic manipulations at the level of DNA, through a variety of back-up systems. This is one of the bases of the robustness of biological systems. Moreover, when we manipulate a gene, e.g. through a mutation, even when phenotypic effects do result they reveal simply the consequences of the difference at the genetic level; they do not reveal all the effects of that gene that are common to both the wild and mutated gene. This is the reason for calling this the 'genetic differential effect problem'. Reverse engineering through modelling at a high level that takes account of all the relevant lower level mechanisms enables us to assign quantitatively the relative roles of the various genes/proteins involved. Thus, a model of pacemaker activity allows absolute quantitative assignment of contributions of different protein transporters to the electric current flows involved in generating the rhythm. Only a few models within the Cardiac Physiome Project are already detailed enough to allow this kind of reverse engineering that succeeds in connecting down to the genetic level, but it must be a goal to achieve this at all levels. This is the reason why top-down analysis, on its own, is insufficient, and is therefore another justification for the middle-out approach.

Modularity in biological systems

Another major principle is that of modularity. A module represents a component of a system that can be relatively cleanly separated from other components. An example is a model of a time- and voltage-dependent ion channel, where the model represents kinetically the behaviour of a large number of identical channel proteins opening more or less synchronously in the same conditions. A model for a cellular action potential would be composed of an assemblage of such modules, each providing the current flow through a different channel type for different ions. Each module is linked to the same environment, but the modules interact with that environment each in their own way. The key to the separability of the modules is that they should be relatively independent of one another, though dependent on their common environment though the effects of each module's behaviour on the environment itself. The separation of modular elements at the same level in the hierarchy works best when the changes in the extramodular environments (concentrations, temperature, pH) do not change too rapidly, that is, more slowly than do the individual channel conductances. The reason is that, when the environmental conditions also change rapidly, the computational 'isolation' of a module becomes less realistic; the kinetic processes represented must extend beyond the module. Choosing the boundaries of modules is important, since a major advantage of modularization is that a limited number of variables are needed to define the interface between models relative to the number required to capture function within the module.

At another level, one might consider the heart, the liver and the lung, etc., as individual modules within a functioning organism, while their common environment (body temperature, blood composition and blood pressure) is relatively stable (homeostasis in Claude Bernard's terms; Bernard, 1865, 1984). At an intermediate level, a module might be composed to represent a part of an organ with a different functional state than other parts, for example, an ischaemic region of myocardium having compromised metabolism and contractile function. Such a module, in an acute phase of coronary flow reduction, might be parametrically identical to the other, normal regions, but have a reduced production of ATP. At a later stage, the regional properties might change, stiffening with increasing collagen deposition, and requiring a different set of equations, so that there would be a substitution for the original module.

In the normal state, a module for any particular region within an organ is inevitably a multiscale

model, containing elements at the protein (channel or enzyme) level, of subcellular regions, of interacting cell groups such as endothelial–smooth muscle–cardiomyocyte arrangements for blood–tissue exchange of nutrients, and with the intracellular responses of each cell type. This level of complexity invites 'model reduction', to save computational time when one has to account for regional heterogeneity within the organ in order to characterize its overall behaviour. No organ yet studied has been found to be homogeneous in all of its functions. Livers, hearts, lungs and brains all exhibit internal heterogeneities; for example, blood flows vary with standard deviations of about 25% of the mean at any particular point in time within an intact healthy organ, and are undoubtedly associated with similar variation in other aspects of their function.

Modules can be envisaged as computational units. Having such units well defined provides for security in archiving, in model sharing and for ease of reproducibility, and for selection in model construction. It also renders those units more accessible and independently modifiable. A given module, e.g. for force generation by muscle contraction, might be cast in several different forms that represent different degrees of fidelity, robustness and biophysical detail. Some versions might be grossly simplified compared with a detailed and thermodynamically correct biophysical/biochemical reference model; such simplified versions could then be used effectively within a multiscale cardiac model for particular physiological states, for example with the onset of cardiac dyssynchrony with left bundle branch block, a situation in which local cardiac contractile work and cardiac glucose metabolism diminish dramatically in the early-activated septal region and increase greatly in the late-activated left ventricular free wall. In this case, the parameters of the metabolic or contractile modules change, but the modules are not necessarily replaced. A principle of modularity is that modules should also be replaceable to allow an appropriate choice for a particular purpose, e.g. when infarction and replacement by scar render the tissue incapable of contracting so that it acts simply as passive elastic material.

Multiscale models are inherently hierarchical; an organ-level module comprises a set of tissue-level modules, and a tissue-level module is composed of a larger set of cell and structural modules. The modules higher in the hierarchy (organ, tissue) are necessarily representing more complex biological functions, so are usually simplified for computation. The result is a loss of the robustness, which lies in the adaptability in cell signalling, protein transcription rates, ATP generation rates, vasomotion, etc. Let us define robustness as the ratio of a perturbing force or demand to the degree of disturbance of the system; an example of strong robustness would be the large change in cardiac output demanded by the body in going from rest to exercise divided by the small change in cytosolic ATP levels in a normal heart. A reduced form module, lacking the cell's metabolic regulatory system, would not be able to respond by increasing its substrate uptake, metabolic reaction rates and ATP production in a finely tuned, automated way.

Technically, module-to-module compatibility requires some standardization in design. In addition to having a name compatible with an ontology, each needs to be identified as to domain and to the inputs and outputs that are needed to communicate with the environment. For an ion channel, the inputs would be the concentrations of solutes on either side of the membrane and the membrane potential. The output would be the current flux as a function of time. The equations for the environmental state (inside and outside the cell) would take the flux and calculate the concentrations and transmembrane potential. The parameters governing the channel conductance can remain hidden from the environment; they are used in computing the conductance as a function of time, but if their values are not needed outside the module, they need not be conveyed, so the information flow is minimized.

Module reduction is presumed desirable; if a governing set of parameters can be held constant then the behaviour of the module (its current flux) is all that needs representation, and the simplest algorithm that does this is adequate. If this is true for a channel, then the same statement can be made for the next level composite module, representing the whole cell excitation–contraction coupling or the whole region of tissue. Successive reductions, each capturing the physiological behaviour of the particular level, can then be made progressively simpler. We might end up, for example, with just a varying elastance representation of each of a set of regions in the heart. While this works in an unchanging physiological state (Bassingthwaighte, 2008) and is useful for limited purposes, there is both a risk and a clear deficiency in the approach. The risk is that the resulting reduced version may be correct over such a small range of physiological variation that the model is incorrect a good fraction of the time, like a stopped clock being correct twice a day. The deficiency is that the model cannot adapt to a change in environment or demand for changes in rate or cardiac output. This difficulty can be partly offset by not taking the reduction quite so far, retaining some links to the subcellular level where adaptations in metabolism, force generation and signal sensing occur; one can even develop sets of alternative, partly reduced model forms, substituting in the version that is most appropriate to the occasion. This begs the next question, how does one automate or use artificial intelligence to make these substitutions or to choose to return to the unreduced, fully detailed model form, as the model is being run. Such automation is critical to

the use of models in diagnostic or clinical monitoring situations.

The criteria for modularity we have outlined above are essentially descriptive criteria, i.e. the criteria to be taken into account when designing modules within computational models. There is a separate question, which is whether the modules we find necessary or convenient in computational models correspond to any modularity displayed by nature. In fact there is no guarantee that nature is organized in modules that correspond to our choices in dividing up the task of simulation. As an example, in the fruit fly, the same gene (the period gene, *per*) may be involved in circadian rhythm, in embryonic development and in modulated wing-beat frequencies used in communication. Many, perhaps most, genes have such multiple functionality, sometimes surprising in their range, like pieces of a child's construction kit that can be re-used to build many different models. Thus, while it may be necessary for us to divide function up according to what we need mathematically and computationally at the higher levels, we should remember that natural reality at the lower levels may more closely resemble tangled forest undergrowth rather than a neatly laid out park.

This may be one of the reasons for the extensive genetic buffering to which we have already referred. The 'mapping' of lower level interactions may not easily correspond to that at higher levels. Many different lower level networks must be capable of subserving a higher level function. Yet connecting low-level genetic and protein network processes to high-level organ function, and the reverse engineering required to use high-level simulation to assess the relative contributions of different genes to overall function (thus solving the 'genetic differential effect problem'; see '*Multiscale analysis*' above) is necessary. It also poses major challenges that have yet to be resolved in the Physiome Project, since the mathematics required at the different levels is usually very different.

A genetic program?

The problem of modularity is related to another deep question in simulating organisms. The discovery of the structure of DNA and of the triplet genetic code almost inevitably led to analogies between organisms and computers. After all, the code itself can be represented digitally, and Monod and Jacob, when they introduced the concept of a genetic program (Monod & Jacob, 1961), specifically noted the analogy with early valve-based digital computers. The DNA corresponded to the tape of instructions and data fed into the machine, while the egg cell corresponded to the machine itself. This analogy also fuelled the concept of organisms as Turing machines. If the genetic code really was a complete 'program of life', readable like the tape of a Turing machine, then it might

follow the Church–Turing computability thesis that every effective computation can be carried out by a Turing machine (Church, 1936; Turing, 1936).

An organism, however, breaks many of the restrictive requirements for a Turing machine. First, information in biological systems is not only digital, it is also analog. Even though the CGAT code within DNA strings can be represented as a digital string, expression levels of individual genes are continuously variable. As we noted earlier, it is the continuous variation in patterns of gene expression that accounts for a heart cell being what it is despite having the same genome as a bone cell. Some computer scientists have argued that analog processing is precisely what is required to go beyond the Turing limit (Siegelmann, 1995). Moreover, gene expression is a stochastic process, displaying large variations (not just experimental noise) even within cells from the same tissue. Such stochasticity is incompatible with determinate Turing-type programming (Kupiec, 2008, 2009).

Second, DNA is not the sole information required to 'program' the organism. Cellular, tissue, organ and system architectures are also involved, including in particular what Cavalier-Smith calls the membranome (Cavalier-Smith, 2000, 2004). Some of these structures (mitochondria, chloroplasts) are now thought to have been incorporated independently of nuclear DNA during evolution by the process known as endosymbiosis (Margulis, 1998). The DNA is therefore not the only 'book of life'. This organelle and other structural information can only be said to be digital in the sense in which we can represent any image (of whatever dimension) digitally within a certain degree of resolution. But, of course, the organism does not use such representation. Finally, there is continuous interaction between genomes and their environment. This interaction can even include environmental and behavioural influences on epigenetic marking of DNA (Weaver *et al.* 2004, 2007). Organisms are therefore interaction machines, not Turing machines (Neuman, 2008).

Nevertheless, the 'genetic program' metaphor has had a powerful effect historically on the way in which we think about modelling life. The idea that we could represent organisms in a fully bottom-up manner is seductive. We suggest that it also underlies the general approach used by many systems biologists, which is to neglect the higher level structural and organizational features. The Physiome Project, in contrast, by including structural and organizational features, provides a mathematical framework for incorporating both genetic and environmental influences on physiological function. In fact, imaging data is central to many of its successes, starting with fully anatomical models of cardiac structure.

While it would be a mistake to reduce organisms to algorithms (in the sense used in the Church–Turing thesis), there is an important role for mathematical

analysis. Brute force computation, however impressive in reproducing biological function and however much computing capacity it may use, is not in itself an explanation. Computation needs to be complemented by mathematical analysis, involving simplifying assumptions to reduce highly complex models to a tractable form to which mathematical analysis might be applied. If we are to unravel the 'logic of life' via the Physiome Project, then such mathematical insights will play an important role.

Mathematical modelling in biology

Before we describe the multiscale modelling framework being developed by the Physiome Project, it is instructive to review the areas in which mathematical modelling is currently being applied to biology and the techniques being used across the huge range of spatial and temporal scales required to understand integrative physiological function from genes to organisms. In order of historical development and using commonly accepted terminology, the major areas of application of mathematical modelling to biology can be summarized as follows.

(1) ***Evolutionary biology and genetics*** is a long-established and sophisticated area of quantitative (i.e. mathematical and statistical, if not model based) analysis that deals with mutation and selection, genotype–phenotype maps, morphogenesis, etc. This field dates from the early 1900s (Pearson, 1898) and is ripe for connection to modern computational physiology as the connection between gene regulatory networks and engineering models of whole organ phenotype is established. In fact, the Physiome Project could be one of the ways in which physiology can reconnect with evolutionary and developmental biology, as advocated, for example, by Jared Diamond (1993).

(2) ***Biophysics and electrophysiology*** comprise a historically distinct modelling area that deals with membrane-bound ion channels, transporters, exchangers and the membranes themselves. This field dates from the 1940s work on neural action potentials that culminated in the Hodgkin–Huxley model (Hodgkin & Huxley, 1952) and for cardiac cells from the 1960s, with the series of papers that began with the first model of a cardiac pacemaker (Noble, 1962). Biological applications that began with giant squid now include many mammalian species. Models are based on combinations of ordinary differential equations, representing membrane capacitative effects and the voltage-dependent gating behaviour of ion channels, and algebraic equations representing biophysical constraints such as charge conservation and the mass conservation of ion species, cast in ordinary and partial differential equations (ODEs and PDEs) and differential algebraic equations (DAEs), the latter being more problematic to solve (Fábián *et al.* 2001).

Recent developments include more sophisticated Markov state models of ion channel gating. This is an advanced field, thanks to 50 years of close interaction between modelling and experiments, especially in the neural and cardiac fields but now also in relation to endothelial cells (Vargas *et al.* 1994), epithelial cells (Thomas *et al.* 2006), pancreatic β-cells (e.g. Sherman & Rinzel, 1991) and links with the more general areas of facilitated membrane transport and cellular metabolism, e.g. in endothelial cells (Bassingthwaighte *et al.* 1989) and hepatic cells (Goresky *et al.* 1973). It is a field that will benefit greatly from the development of coarse-grained approximations of molecular dynamics/quantum mechanics simulations of ion channel proteins, as the three-dimensional structure for these becomes available from X-ray diffraction or nuclear magnetic resonance imaging or, in the future, predictable from sequence.

(3) ***Mathematical biology*** is a more general field dating from the 1960s, based primarily on ODEs, DAEs and PDEs (primarily reaction–diffusion systems). Areas of study include cancer modelling, the cell cycle and pattern formation in embryogenesis. The link with experiments is less developed than in the electrophysiology field and the link to biophysical mechanisms has also not been a priority, possibly because the practitioners are mathematicians working in isolation from physiologists, although this is now beginning to change. There have been some attempts to link this field to clinical issues, such as the work on diagnosing ECG patterns with low-dimensional state space dynamics (Peng *et al.* 1993).

(4) ***Computational physiology*** is the study of structure–function relations at the levels of cells, tissues and organs using anatomically and biophysically based models. The field, which has traditionally had a strong clinical focus, dates from the 1960s, when it was confined to tissues and organs and was dominated by engineers studying stress analyses in bone and the influence of fluid mechanics on arteriosclerosis. Biomechanics is now a mature field, particularly in its application to musculo-skeletal mechanics and to arterial flow mechanics. The multiscale version of computational physiology is now encompassed in the Physiome Project and deals both with the interaction of different types of biophysical equation at the tissue level (reaction–diffusion equations, large deformation solid mechanics, fluid mechanics, etc.) and the connection with models of subcellular function (ion channels, myofilament mechanics, signal transduction, metabolic pathways, gene regulation, etc.). The mathematical techniques derive from engineering physics (conservation laws, continuum mechanics, finite element methods, etc., coupled to DAEs at the cell level) and are heavily dependent on high-performance computation (Hunter *et al.* 2003). It is a field that is also increasingly dependent on computer science (markup languages, ontologies, the Semantic Web, etc.) in order to cope with the explosion of complexity that

occurs when tissue modelling is extended down to protein pathways and associated bioinformatic databases (Hunter *et al.* 2008). An important direction for this field is the use of anatomically based models at the cell level. One example of how bioengineering approaches are contributing to our understanding of structure–function relations at the cell level is the cellular biomechanics area, which has successfully incorporated molecular mechanisms into a continuum mechanics framework (Kamm & Mofrad, 2006).

(5) ***Computational chemistry*** (as applied to biology) is the analysis of protein–protein and protein–ligand interactions using three-dimensional atomic structure. Quantum mechanics is used to a limited extent, but the dominant method is molecular dynamics (MD), which applies $F = ma$ to the atomic masses, where F is the combination of covalent bonds, van der Waal forces and electrostatic interactions, m is mass and a is acceleration. A major challenge for this area is that biological models require hundreds of thousands of atoms, so it is not computationally feasible to run them for more than a few nanoseconds of real time (equivalent to months of high-performance computer time). This is enough to compute the permeability of an aquaporin water channel but is not enough to predict the gating behaviour of an ion channel. There is some (not enough) work on coarse-grained approximations of MD. This is needed in order to derive the ion channel gating, possibly via the parameters of Markov gating models, from the atomic structure of ion channels. Hopefully, at some point this field will also allow us to predict protein folding from residue sequence.

(6) ***Network systems biology*** is the analysis of networks, primarily gene regulatory networks, signal transduction pathways and metabolic networks. The field has been strongly promoted by Hiroaki Kitano (Kitano, 2002) and Leroy Hood (Facciotti *et al.* 2004) over the last 10 years and has benefitted greatly from the development of the Systems Biology Mark-up Language (SBML) standard and a rich set of software tools coupled with the SBML-encoded BioModels Database (www.ebi.ac.uk/biomodels-main). Yeast is the common biological species studied by systems biologists, because it is one of the simplest eukaryotic organisms. The equations are DAEs, and the network properties of interest are, for example, modularity, transient/oscillatory responses, state space attractors and robustness to stochastic perturbation. The equations include biophysics only to the extent of the mass balance of chemical species (and usually including Michaelis–Menten enzyme-catalysed reaction kinetics). Mathematical analysis is based on state space techniques (e.g. low dimensional dynamics, bifurcation theory, etc.), non-linear control theory and Bayesian statistical methods, as well as a whole range of classical linear algebra techniques such as singular value decomposition, dynamic programming, etc.

(7) ***Systems physiology*** is the study of integrated physiological function at the organ system level using DAE models. The field was initiated with the blood pressure control work of Guyton in the 1960s (Guyton *et al.* 1972) but has stalled. It is ripe for a renaissance based on modern computational techniques and the use of the CellML and FieldML encoding standards. For example, one-dimensional bioengineering models of the cardiovascular system and the lymphatic system could be used with cellular models of exocrine signalling and lumped parameter versions of organs that are linked through multiscale techniques to computational physiology models. Similarly, we need to connect the field of clinical musculo-skeletal biomechanics to computational physiology models of the three-dimensional structure of bones, muscles, tendons, ligaments and cartilage.

Physiome Project infrastructure

To address the challenges of multiphysics and multiscale modelling in computational physiology, the Physiome Project is developing modelling standards, model repositories and modelling tools. The key elements of this modelling infrastructure are CellML and its related model repositories and tools, and FieldML and its related model repositories and tools.

CellML (www.cellml.org) is an XML markup language developed to encode models based on systems of ODEs and DAEs. CellML deals with the structure of a model and its mathematical expression (using the MathML standard) and also contains additional information about the model in the form of metadata; such things as: (1) bibliographic information about the journal publication in which the model is described; (2) annotation of model components in order to link them to biological terms and concepts defined by bio-ontologies such as GO (the gene ontology project); (3) simulation metadata to encode parameters for use in the numerical solution of the equations; (4) graphing metadata to specify how the output of the model should be described; and (5) information about the curation status of the model (Cooling *et al.* 2008). Once a model is encoded in CellML, the mathematical equations can be automatically rendered in presentation MathML or can be converted into a low-level computer language such as C, C++, Fortran, Java or Matlab. A number of simulation tools are available to run CellML models, for example, PCEnv (www.cellml.org/tools/pcenv/), COR (cor.physiol.ox.ac.uk/), JSim (www.physiome.org/jsim/) and Virtual Cell (www.vcell.org/).

Note that another similar markup language, SBML (www.sbml.org), has also been developed for models of gene regulation, protein signalling pathways and

metabolic networks. This language has widespread acceptance in the systems biology community, and many systems biology analysis tools have been developed that are compatible with SBML. CellML has a broader but complementary scope in that it deals with biophysically based models, whereas SBML is focused on biochemical networks. The CellML and SBML developers frequently exchange ideas, and models can fairly readily be converted between the two formats. At some point it may be advantageous to merge the two languages.

FieldML (www.fieldml.org) is an XML markup language being developed to work in conjunction with CellML for encoding spatially varying and time-varying fields within regions of an organ or tissue. The language is designed to support the definition and sharing of models of biological processes by including information about model structure (how the parts of a model are organizationally related to one another), mathematics (such as 2-dimensional and 3-dimensional partial differential equations describing concentrations or other variables over the fields) and metadata (additional information about the model). FieldML will describe spatially varying quantities such as the geometric co-ordinates and structure of an anatomical object, or the variation of a dependent variable field such as temperature or oxygen concentration over that anatomical region.

These standards are now being applied across a very wide range of physiological function. The model repositories for CellML (www.cellml.org/models), Bio-Models (www.biomodels.org), the National Simulation Resource Physiome site (www.physiome.org/Models), and JWSmodels (http://jjj.biochem.sun.ac.za) cover most aspects of cellular function and many areas of organ system physiology, albeit to varying degrees, across the body's twelve organ systems. The recently established European Network of Excellence for the Virtual Physiological Human (VPH; www.vph-noe.eu/index.php) is providing a major boost to the development of the standards, model repositories and tools and in particular their clinical applications.

Another integrative effort was initiated within the US federal science support system. The National Institute of General Medical Sciences began issuing requests for applications in support of integrative biology and modelling in 1998. In April of 2003, an Interagency Modeling and Analysis Group (IMAG) was formed, starting from a working group comprised of program staff from nine Institutes of the National Institutes of Health (NIH) and three directorates of the National Science Foundation (NSF). The IMAG now represents 17 NIH components, four NSF directorates, two Department of Energy (DOE) components, five Department of Defense (DOD) components, the National Aeronautics and Space Administration (NASA), the United States Department of Agriculture (USDA) and the United States

Department of Veterans Administration (USDVA) (see www.nibib.nih.gov/Research/MultiScaleModeling/IMAG). Since its creation, this group has convened with monthly meetings at various locations of the IMAG agency participants, and less frequent meetings of the investigators from the 30 funded projects, often as phone or Web presentation/discussions, and annually as workshops. The ten IMAG investigator-led Working Groups develop collaborations in technologies and in science, and share models and technologies.

Clinical applications

The intention of the Physiome Project is to span medical science and its applications 'From Genes to Health', the title of a 1997 Coldspring Harbor symposium and the central theme of the 1997 Physiome IUPS Satellite meeting near St Petersburg. Clinical applications include clinical image interpretation using positron emission tomography (PET), magnetic resonance imaging (MRI) and ultrasound. The interpretation of cardiac PET image sequences, for example, requires models for blood–tissue exchange and metabolic processes using PDEs and ODEs. The result is the production of 'functional images' displaying cardiac three-dimensional maps of flow, metabolic rates of utilization of oxygen or glucose or other substrates (e.g. thymidine in tumours), and the regional densities of receptors in people with arrhythmia or cardiac failure (Caldwell *et al.* 1990, 1998; Wilke *et al.* 1995). Current work in progress concerns the development of automated algorithms for image capture, segmentation and region-of-interest selection, estimation of the particular function through optimization of model fits to the data and construction of the three-dimensional functional image and report.

Where to now?

Over the last 10 years, the IUPS Physiome Project has focused more on computational physiology, while the NSR Physiome group has worked on clinical and research applications and teaching models. These groups and others have been developing computational infrastructure for combining models at the cell, tissue, organ and organ system levels. The Physiome Project is now well underway, with the markup languages, model repositories and modelling tools advancing rapidly. But this is only the very beginning, and there are many challenges ahead. One challenge is the connection to networks systems biology, which should be relatively straightforward because the field has already widely adopted the SBML standard that is closely related to the Physiome standard, CellML. Another challenge is to link the Physiome models to clinical data in a broader way than as described under '*Clinical applications*' and to link the models to the standard image

formats, such as DICOM (http://medical.nema.org/), that are widely used in clinical imaging devices such as MRI and computed tomography (CT) scanners. The use of FieldML should help greatly with this, and discussions are now underway to include FieldML files within the public or private header tags of a DICOM file in order to include the fitted parametric model with the clinical images used in its creation. A related goal is to use models for bedside assessment of clinical status, requiring real-time simulation and data analysis to provide (or suppress) alarms and to guide therapy (Neal & Bassingthwaighte, 2007). A particular computational problem is to link three-dimensional finite-element and finite-volume models to one- or two-dimensional models that can provide the boundary conditions; this technology is needed in order to create the setting for the proper physiological behaviour of the three-dimensional models, since they are far too computationally intensive to be used for a whole system at once. Another challenge is to bring the computational physiology models of the Physiome Project to bear on the fascinating genotype–phenotype questions that have occupied the minds of evolutionary biologists for over 100 years. As we noted earlier in this article (see '*Multiscale analysis*'), high-level phenotype modelling can in principle solve a major problem in genotype–phenotype relations, i.e. what we have called the 'genetic differential effect problem'.

To date, most of the insights gained into physiological processes from mathematical models have been derived from models that deal with one or more physical processes but at only one spatial scale. Examples are models of mechanical processes in the heart, gas transport in the lungs, arterial blood flow dynamics and lipid uptake into endothelial cells, and stress–strain analysis to assist with prosthetic implant design in musculo-skeletal joints. There are a few examples of multiscale analysis, for example, models of electrical activation waves in the heart that are linked to ion channel kinetics. There is a pressing need to be doing much more multiscale analysis. For example, the most important aspect of joint implant design is how to avoid the bone remodelling that leads to implant loosening, and this problem can only be tackled by linking tissue-level stress analysis to protein-level cell signalling pathways. A similar requirement exists in the cardiac mechanics field, where the processes underlying heart failure are governed by a combination of tissue-level stress from raised blood pressure (for example) and gene regulatory processes that alter the protein composition of the tissue.

Probably the biggest challenge facing the Physiome Project now, and one that is crucial to the computational feasibility of multiscale analysis, is that of model reduction. Automated methods are needed to analyse a complex model defined at a particular spatial and temporal scale in order to compute the parameters of a simpler model that captures the model behaviour relevant to the scales above. For example, if the three-dimensional atomic structure of an ion channel is known, a molecular dynamics model can be formed to compute channel conductance, but one would like to compute the parameters of a much simpler (Hodgkin–Huxley or Markov state) model of the current–voltage channel phenotype appropriate for understanding its behaviour at the cell level. Similarly, it would be highly desirable to be able to derive a model of multichannel cell-level action potential phenotype that accounted for current load from surrounding cells and could be used efficiently in larger scale models of myocardial activation patterns in the whole heart.

The mathematical challenges of deriving automated model reduction methods are greatly facilitated by the model encoding standards that have been put in place over the last 10 years.

Conclusion

We are optimistic that within the next 10 years multiscale analysis based on automated model reduction will be a well-honed tool in the hands of physiologists and bioengineers. Over that time scale, we can expect a significant number of important applications of the Cardiac Physiome Project in the healthcare field.

What of the prospects of more fundamental contributions to the conceptual foundations of biology, i.e. the questions with which we began this article? It is inherently hazardous to predict the development of concepts. If we could, they wouldn't be predictions. But we hope that by drawing attention to these issues and indicating how the Physiome Project may also contribute to the conceptual foundations of biology, we will have encouraged adventurous physiologists, mathematicians, engineers and computer scientists to tackle those problems.

References

Bassingthwaighte JB (2008). Linking cellular energetics to local flow regulation in the heart. *Ann N Y Acad Sci* **1123**, 126–133.

Bassingthwaighte JB, Wang CY & Chan I (1989). Blood-tissue exchange via transport and transformation by endothelial cells. *Circ Res* **65**, 997–1020.

Bernard C (1865, 1984). *Introduction à l'étude de la médecine expérimentale*. Flammarion, Paris.

Bird A (2007). Perceptions of epigenetics. *Nature* **447**, 396–398.

Caldwell JH, Kroll K, Seymour KA, Link JM & Krohn KA (1998). Quantitation of presynaptic cardiac sympathetic function with carbon-11-meta-hydroxyephedrine. *J Nucl Med* **39**, 1327–1334.

Caldwell JH, Martin GV, Link JM, Gronka M, Krohn KA & Bassingthwaighte JB (1990). Iodophenylpentadecanoic acid–myocardial blood flow relationship during maximal exercise with coronary occlusion. *J Nucl Med* **30**, 99–105.

Cavalier-Smith T (2000). Membrane heredity and early choroplast evolution. *Trends Plant Sci* **5**, 174–182.

Cavalier-Smith T (2004). The membranome and membrane heredity in development and evolution. In *Organelles, Genomes and Eukaryite Phylogeny: An Evolutionary Synthesis in the Age of Genomics*, ed. Hirt RP & Horner DS. CRC Press, Boca Raton.

Church, A (1936). An unsolvable problem of elementary number theory. *Am J Math* **58**, 345–363.

Cooling MT, Hunter PJ & Crampin EJ (2008). Modeling biological modularity with CellML. *IET Syst Biol* **2**, 73–79.

Diamond JM (1993). Evolutionary physiology. In *The Logic of Life*, ed. Boyd CAR & Noble D. Oxford University Press, Oxford.

Fábián G, van Beek DA & Rooda JE (2001). Index reduction and discontinuity handling using substitute equations. *Math Comput Model Dynam Syst* **7**, 173–187.

Facciotti MT, Bonneau R, Hood L & Baliga NS (2004). Systems biology experimental design – considerations for building predictive gene regulatory network models for prokaryotic systems. *Curr Genomics* **5**, 527–544.

Goresky CA, Bach GG & Nadeau BE (1973). On the uptake of materials by the intact liver: the transport and net removal of galactose. *J Clin Invest* **52**, 991–1009.

Guyton AC, Coleman TG & Granger HJ (1972). Circulation: overall regulation. *Annu Rev Physiol* **34**, 13–46.

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.

Hunter PJ, Crampin EJ & Nielsen PMF (2008). Bioinformatics, multiscale modeling and the IUPS Physiome Project. *Brief Bioinform* **9**, 333–343.

Hunter PJ, Pullan AJ & Smaill BH (2003). Modelling total heart function. *Rev Biomed Eng* **5**, 147–177.

Kamm RD & Mofrad MK (2006). *Cytoskeletal Mechanics: Models and Measurements*. Cambridge University Press, Cambridge, UK.

Kitano H (2002). Systems biology: towards systems-level understanding of biological systems. In *Foundations of Systems Biology*, ed. Kitano H. MIT Press, Cambridge, MA, USA.

Kupiec J-J (2008). *L'origine des individus*. Fayard, Paris.

Kupiec J-J (2009). *The Origin of Individuals: A Darwinian Approach to Developmental Biology*. World Scientific Publishing Company, London.

Margulis L (1998). *Symbiotic Planet: A New Look at Evolution*. Basic Books, New York.

Monod J & Jacob F (1961). Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb Symp Quant Biol* **26**, 389–401.

Neal ML & Bassingthwaighte JB (2007). Subject-specific model estimation of cardiac output and blood volume during hemorrhage. *Cardiovasc Eng* **7**, 97–120.

Neuman Y (2008). *Reviving the Living: Meaning Making in Living Systems*. Elsevier, Amsterdam and Oxford.

Noble D (1962). A modification of the Hodgkin–Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *J Physiol* **160**, 317–352.

Noble D (2006). *The Music of Life*. OUP, Oxford.

Noble D (2008*a*). Claude Bernard, the first systems biologist, and the future of physiology. *Exp Physiol* **93**, 16–26.

Noble D (2008*b*). Genes and causation. *Philos Transact A Math Phys Eng Sci* **366**, 3001–3015.

Pearson K (1898). Mathematical contributions to the theory of evolution. On the law of ancestral heredity. *Proc R Soc B Biol Sci* **62**, 386–412.

Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE & Goldberger AL (1993). Long-range anticorrelations and non-Gaussian behavior of the heartbeat. *Phys Rev Lett* **70**, 1343–1346.

Sherman A & Rinzel J (1991). Model for synchronization of pancreatic β-cells by gap junction coupling. *Biophys J* **59**, 547–559.

Siegelmann HT (1995). Computation beyond the Turing limit. *Science* **268**, 545–548.

Thomas SR, Layton AT, Layton HE & Moore LC (2006). Kidney modelling: status and perspectives. *Proc IEEE* **94**, 740–752.

Turing AM (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proc Lond Math Soc 2* **42**, 230–265.

Vargas FF, Caviedes PF & Grant DS (1994). Electrophysiological characteristics of cultured human umbilical vein endothelial cells. *Microvasc Res* **47**, 153–165.

Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Sekl JR, Dymov S, Szyf M & Meaney MJ (2004). Epigenetic programming by maternal behavior. *Nat Neurosci* **7**, 847–854.

Weaver ICG, D'Alessio AC, Brown SE, Hellstrom IC, Dymov S, Sharma S, Szyf M & Meaney MJ (2007). The transcription factor nerve growth factor-inducible protein A mediates epigenetic programming: altering epigenetic marks by immediate-early genes. *J Neurosci* **27**, 1756–1768.

Wilke N, Kroll K, Merkle H, Wang Y, Ishibashi Y, Xu Y, Zhang J, Jerosch-Herold M, Mühler A, Stillman AE, Bassingthwaighte JB, Bache R & Ugurbil K (1995). Regional myocardial blood volume and flow: first-pass MR imaging with polylysine-Gd-DTPA. *J Magn Res Imaging* **5**, 227–237.