The Journal of Experimental Biology 210, 1576-1583
Published by The Company of Biologists 2007
doi:10.1242/jeb.000133

Computational biology of cardiac myocytes: proposed standards for the physiome

Nicolas P. Smith[1,2,*], Edmund J. Crampin[2], Steven A. Niederer[1,2], James B. Bassingthwaighte[3] and Daniel A. Beard[4]

[1]*University Computing Laboratory, University of Oxford, Oxford, OX1 3QD, UK,* [2]*Bioengineering Institute, University of Auckland, Auckland, New Zealand,* [3]*University of Washington, Seattle, USA and* [4]*Medical College Wisconsin, WI, USA*

*Author for correspondence (e-mail: nic.smith@comlab.ox.ac.uk)

Accepted 19 December 2006

Summary

Predicting information about human physiology and pathophysiology from genomic data is a compelling, but unfulfilled goal of post-genomic biology. This is the aim of the so-called Physiome Project and is, undeniably, an ambitious goal. Yet if we can exploit even a small proportion of the rich and varied experimental data currently available, significant insights into clinically important aspects of human physiology will follow. To achieve this requires the integration of data from disparate sources into a common framework. Extrapolation of available data across species, laboratory techniques and conditions requires a quantitative approach. Mathematical models allow us to integrate molecular information into cellular, tissue and organ-level, and ultimately clinically relevant scales. In this paper we argue that biophysically detailed computational modelling provides the essential tool for this process and, furthermore, that an appropriate framework for annotating, databasing and critiquing these models will be essential for the development of integrative computational biology.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1576/DC1

Key words: physiome, mathematical modelling, cardiac, multi-scale.

Introduction

Coupling genomic data to physiological function is the aim of biology in the post-genomic era. Quantitative descriptions of biological processes using mathematical modelling are one important tool in this aim. Empirically derived relationships have commonly been used in modelling biological processes. While such data-driven models often give useful descriptions and insights into specific data sets, these models often fail when combined together to study interaction between multiple processes. On the other hand, physics-based models – models built on principles including the laws of mechanics and thermodynamics, in which assumptions and approximations are made explicit – operate with a common currency of mass, charge, energy and momentum (Bassingthwaighte et al., 2001; Qian et al., 2003). With care, such models may be naturally integrated together to form comprehensive models of biological systems.

It is our conviction that a high degree of the true complexity of the biological mechanisms must be represented in models if clinically applicable insights are to be gained from model simulations. There are, however, significant challenges to be overcome, both mathematical and computational. Multi-scale models must incorporate nontrivial biological complexity, while remaining computationally tractable. Furthermore, while representing this complexity, models must still be capable of providing insights *via* mathematical analysis when simulations do not behave as expected (as must sometimes happen if we are to learn anything new!). This requires the development of approaches to deal with model complexity and parameterization, and communication and information sharing between developers of models.

One approach to handling complexity across multiple spatial and temporal scales is to adopt a modular and hierarchical approach to modelling biological systems. In this approach, mathematical representations of biological components are brought together and tuned appropriately to produce a model of a specific cell or tissue type. The most transparent way of achieving this goal is to retain biophysical detail at each level in a modelling hierarchy, while employing simplifying assumptions to move to higher level descriptions (Smith et al., 2004; Smith et al., 2000). This often requires the coupling of models governed by different physical equations, representing

physiologically discrete functions (Nickerson et al., 2005). Such a hierarchical and multi-physics approach provides an obvious mechanism for revision or improvement of selected parts of a large-scale simulation as new data are collected. Furthermore, this biophysical approach provides greater confidence in the ability of a model to extrapolate from the data used for parameterization and to provide detailed, even patient-specific, predictions when data from an individual are available.

The integration of biophysically based models covering the breadth of physiological function, across spatial and temporal scales, is the approach and philosophy driving the IUPS sponsored Physiome Project (Crampin et al., 2004; Hunter and Borg, 2003). As part of this umbrella project, this multiscale modelling approach has had demonstrable success in models including the gastro-intestinal (Buist et al., 2006), renal (Ribba et al., 2006) and musculo-skeletal organ systems (Hunter et al., 2005) and, arguably the most sophisticated exemplar, the heart or 'cardiome' (Hunter and Borg, 2003). It is from this cardiac work that we draw our examples below; however, the principles we illustrate are relevant across the full range of organ systems.

Typically, as our knowledge and understanding of biological processes grows, models of increasing detail and comprehensiveness have been developed, often by piecing together existing model components, in order to incorporate more and more of the available data. However, the strength of building on existing work can also be the greatest weakness of this approach. Errors and implicit assumptions contained in foundation elements of models can, as we will demonstrate below, propagate through as more complete models are developed. It is, therefore, vital that the assumptions used to develop models are made explicit, and that propagation of errors is prevented. This imposes an extremely high duty of care on both authors and reviewers of new models. In particular, it is unreasonable to expect such problems to come to light during the conventional reviewing process. We assert that new and innovative processes and criteria must be developed to augment the standard peer review process, such that, not only are errors in models eliminated, but also the conditions of appropriate model use and connection with the experimental data are made transparent for the user community. If these issues can be addressed, we believe the scientific community at large will have improved confidence in the fidelity of individual models, and the utility of computational biology as a whole. This will be essential for computational modelling to achieve its promise, both in the laboratory and in the clinic.

Work in a number of groups is already progressing towards the development of tools and ontologies (Cuellar et al., 2003; Schilstra et al., 2006) to facilitate the unambiguous machine-readable representation of biological models. Most recently this concept has been progressed further with the proposal of set of rules (termed MIRIAM, Minimum Information Requested In the Annotation of biochemical Models) for curating quantitative models of biological systems (Le Novere et al., 2005). This community effort defines procedures for encoding and annotating models represented in machine-readable form which, if adopted, should ensure (i) consistency between curated models and their reference description; (ii) provide searchable databases of models using biological terms from accepted ontologies; and (iii) facilitate model reuse and development in the manner that we have described. These rules for annotation do not, however, provide any comment on the nature of the models themselves, or their suitability for any specific modelling purpose (indeed, this is not the intention of the MIRIAM initiative); however, it is apparent that additional constraints on the structure of models will also be useful when combining them together. Below, we briefly review the development of cardiac models with a more detailed focus on four of our own published models. We then highlight two specific examples in the cardiac field where reuse of elements has led to the connection between model parameters and experimental measurement becoming disconnected. These examples are used to motivate the proposal of additional criteria for biophysically based models to address the issues discussed above, before specifically analysing our four published models against these proposed criteria.

The development of integrated cardiac models

The last 40 years have seen the development of increasingly detailed biophysically based cell models of cardiac electrophysiology (Luo and Rudy, 1991; McCulloch et al., 1998). These models currently provide detailed representations of membrane-bound channels and transporters, and fluxes of ions between the cytosol and intracellular organelles. One example of a transporter model is our recent study characterising the kinetics of the sodium pump (Smith and Crampin, 2004) (Fig. 1A). The function of this exchanger is the maintenance of both the sodium and potassium gradients across the myocyte membrane. The kinetics of this process were represented using an enzymatic cycle, formulated to be thermodynamically consistent in coupling the free energy of ATP hydrolysis to movement of the ions against their electrochemical gradients, and fitted to experimental data of observed pump cycling rates at different extracellular sodium concentrations.

The known details of channels, pumps and exchangers have enabled analysis of the role that each functional element plays in health and disease (Shaw and Rudy, 1997). Further, they have provided a successful paradigm for integrating individual data sets on the different molecular components of the cell into a common framework. This allows trans-membrane ion transport to be linked to action potential recordings, in altered ionic conditions, in the whole myocyte, across a range of species from rat to human (Pandit et al., 2001; ten Tusscher et al., 2004). We recently published a model (shown schematically in Fig. 1B) of the myocyte that builds on the existing Luo–Rudy dynamic (LRd) electrophysiology model (Hund and Rudy, 2004). The LRd model was developed to study myocyte electrophysiology over one heart beat. Our

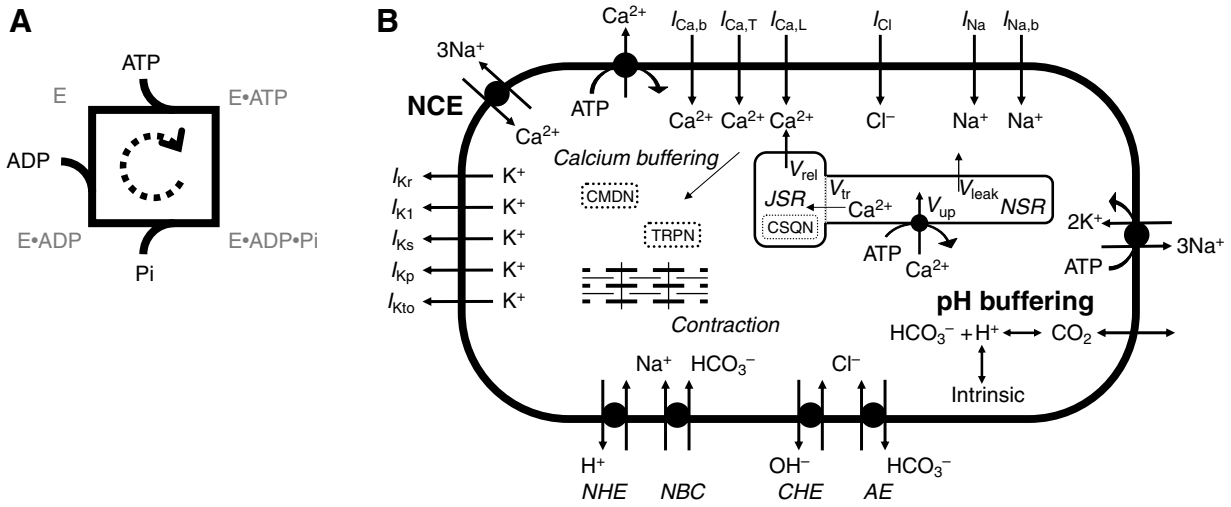

Fig. 1. Schematic diagrams of (A) a 'thermodynamic box' representation of an enzymatic cycle coupled to ATP hydrolysis, as used to model the sodium pump, and (B) the electrophysiological, contraction and pH regulatory components in the coupled myocyte model. E, enzyme; *I*, current. Other abbreviations and further explanation are available elsewhere (Crampin and Smith, 2006).

study considered the effect of acidosis (a drop in pH associated with impaired metabolism) on excitation–contraction coupling in the heart cell, over multiple beats (Crampin and Smith, 2006). This imposes a new set of requirements on the model. It was necessary to ensure conservation of mass and charge, and that under normal conditions the time courses for state variables (ionic concentrations and membrane potential) were maintained from one beat to the next. Our model uses thermodynamically constrained cycles to represent acid transporters and includes proton inhibition of many of the calcium-handling process in the cell, fitted from available experimental data.

While initially lagging behind developments in electrophysiology, cellular models of myocardial contraction have now progressed so that myocardial mechanics can be computationally simulated. Detailed Ca^{2+}-induced activation of thin-filament kinetics has been combined with a representation of cross-bridge tension generation, which describes the length and tension-dependent Ca^{2+}-induced activation of cellular contraction. Transient Ca^{2+}-induced excitation–contraction has been characterized by coupling electrophysiological and mechanical models (Nickerson et al., 2001), thus enabling simulations of activation-induced contraction. Based on the existing framework of Hunter et al. (Hunter et al., 1998), we recently developed a model of active contraction of the myocyte, which uses mass-action kinetics to model calcium binding to TnC, and tropomyosin kinetics (Niederer et al., 2006). These elements have been combined with a phenomenological representation of actin–myosin binding kinetics and the force and length dependence of each process was characterized in detail. In this study, each parameter was rationalized from numerous sources and, where possible, multiple experimental modalities, through an extensive review of the literature (Fig. 2A is shown as an example). Issues of species consistency and experimental conditions, in particular temperature, are explicitly addressed in the choice of parameters to represent a rat myocyte at room temperature.

In parallel work, we have developed a computational model of muscle cell oxidative energy metabolism (illustrated in Fig. 2B), which we have applied to analyze cardiac and skeletal muscle energetics (Bassingthwaighte et al., 2001; Wu et al., 2007). In these studies, ATP consumption is treated as a forcing function and the ATP consuming processes associated with contraction and electrophysiology are not explicitly modelled. In current work, the energy metabolism model is being integrated with the electrophysiology and mechanics models, leading to an increasingly detailed model of cardiomyocyte biophysics.

Despite the increasing complexity, rapid improvements in the performance per unit cost of high performance computing has more than offset the computational demands for solving the systems of ordinary differential equations that represent these cellular and sub-cellular models. This has led to the development of models of cardiac tissue, in which the cellular models are embedded in a continuum description of tissue geometry. These models incorporate data from confocal microscopy, which detail the myocyte, fibroblast and collagen microstructure within the tissue. These microstructural data can be used to determine the conductivity and stiffness tensor within the continuum model, in order to predict the functional properties of electrical conductivity and mechanical stiffness of cardiac tissue (Trew et al., 2006). By applying the mono-domain or bi-domain equations, tissue-level models have been used to predict the spread of activation in two- and three-dimensional simulations (Smith et al., 2004; Tomlinson et al., 2002). Using the tension transients calculated in the cellular models, tissue deformation can be predicted by solving the equations of finite deformation (Pullan et al., 2001). Linking the calcium transient of the cellular electrophysiology model to cellular tension generation enables the coupling of activation

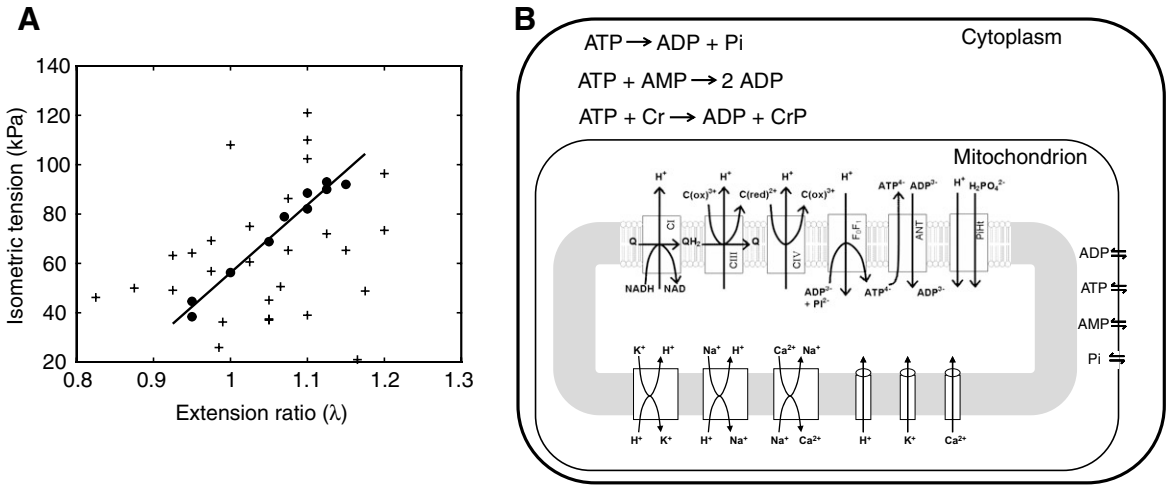

Fig. 2. (A) Isometric tension data at varying strains. Solid data points represent measurements taken under physiological conditions used to fit the model. The remaining data are plotted as crosses. (B) Schematic diagram of the model of muscle cell oxidative energy metabolism.

and contraction. This coupling is achieved at the tissue level by combining numerical solution techniques properly to preserve computational efficiency (Nickerson et al., 2005; Smith et al., 2003) (Fig. 3).

In this way, cellular and sub-cellular modelling provides a framework for capturing mechanisms at their own spatial scale and for extrapolating these responses to determine behaviour at the tissue level. The parameters of each of these cellular models are typically determined either directly (a single measurable parameter) or indirectly (fitting a data set) from experimental data.

It is critical to preserve this link to experimental data, both for appropriate parameterisation and for validation of model function. The potential provided by the ability to reuse and integrate existing model components can, however, be a double-edged sword. Model integration leads to the reuse of parameters, which is a necessary and efficient means to generate new, more complex models. Even if all model parameters are determined using the best currently available experimental data, they may still be superseded in time. The parameter set for a model component can, however, become obscured from further reviewer scrutiny once it is reused in later models, and the original explicit connection with experimental data is lost.

Specific cases of this phenomena for the propagation of two common cardiac myoctye model parameters over 25–30 years of modelling are shown in Fig. 4A,B: the binding affinity of Ca^{2+} to troponin C (Crampin and Smith, 2006; Faber and Rudy, 2000; Hilgemann and Noble, 1987; Holroyde et al., 1980; Hunter et al., 1998; Jafri et al., 1998; Luo and Rudy, 1994; Nickerson et al., 2001; Noble et al., 1998; Pandit et al., 2001; Robertson et al., 1981; Rodriguez et al., 2002; Winslow et al., 1999; Zeng et al., 1995) and to calsequestrin (Bondarenko et al., 2004; Cannell and Allen, 1984; Crampin and Smith, 2006; Faber and Rudy, 2000; Hund and Rudy, 2004; Iyer et al., 2004; Jafri et al., 1998; Luo and Rudy, 1994; Ostwald and MacLennan, 1974; Pandit et al., 2001; ten Tusscher et al., 2004; Winslow et al., 1999; Zeng et al., 1995). In both cases, an early model (Cannell and Allen, 1984; Robertson et al., 1981) provided a foundation component for a number of the current cardiac models. Since the original models were published, there has been a consistent flow of new and arguably more reliable experimental data sets, which have been largely ignored by the modelling community. The vast majority of cardiac models (including our own) (Crampin and Smith, 2006) are guilty of building on existing models without considering the source of all the model parameters. To address this issue, in our recent model of active contraction (Niederer et al., 2006) we performed an extensive literature search for each model parameter and noted the experimental conditions under which

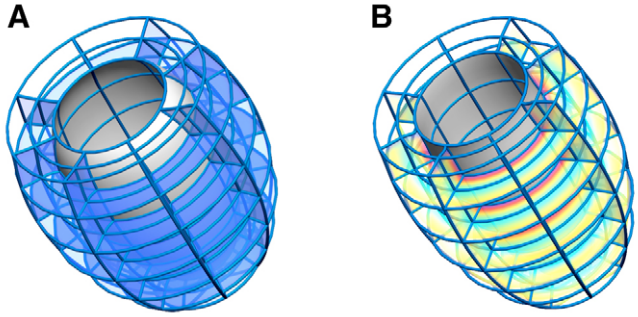

Fig. 3. Coupled electromechanics simulation at diastolic (A) and systolic (B) states. The coloured surfaces indicate active tension with blue corresponding to 0 kPa and red to 50 kPa. The model uses a simplified left ventricular geometry, tension is calculated using the electrophysiology model (Crampin and Smith, 2006) coupled with the active contraction model (Niederer et al., 2006), and passive material laws are defined by the Pole Zero law (Nash and Hunter, 2000). The equations were solved as previously described (Nickerson et al., 2005).

the parameter was measured. We belive this adoption of clear links between model parameters and experimental results is an important step in maintaining credibility in cardiac modelling.

Criteria for model assessment

Systematic validation against experimental data of models linking detailed cellular biophysics to tissue function remains challenging. As outlined above this is, in part, due to the technical difficulties associated with managing and maintaining links to experimental data required for each mechanism in the excitation–contraction metabolism process. Nonetheless, validation is essential before these promising simulation techniques can provide real value to the clinician.

The specific difficulties outlined above are as follows. (1) Models are rarely implemented and tested as part of the peer-review process for journal publications, meaning the published manuscript may contain errors. (2) The connection between model parameters and data is often ambiguous. Making this link transparent is fundamental to building large-scale models that integrate different physiological subsystems. (3) The functional limitations of a model do not become apparent until significant time and effort has been put into model implementation, application and coupling. (4) There are few public forums where feedback, experiences and critique of existing published models can be shared. (5) The experimental data used to parameterize and validate computational models are rarely available to the community in convenient useable formats.

Each of these issues undermines confidence and impairs the application and extension of models by people other than the developers, or those with specific expertise in model development. As discussed above, a number of cell modelling mark-up languages have been developed (CellML, SBML, Jsim) and using these, and other established computing languages, cell models can be made freely available. Furthermore, there is on-going discussion of the development of FieldML (http://www.physiome.org.nz/fieldml/pages/), a mark-up language that will enable the representation of structural and continuum information about biological and physical entities. This will allow the unambiguous machine-readable representation of structural and tissue-based models. Running versions of models provided by model authors using these codes provides a significant step in overcoming issue 1. Furthermore, a model that is compliant against the MIRIAM rules guarantees machine readability, an unambiguous description of the model, consistency with the published model, and consistency between published results and simulation output.

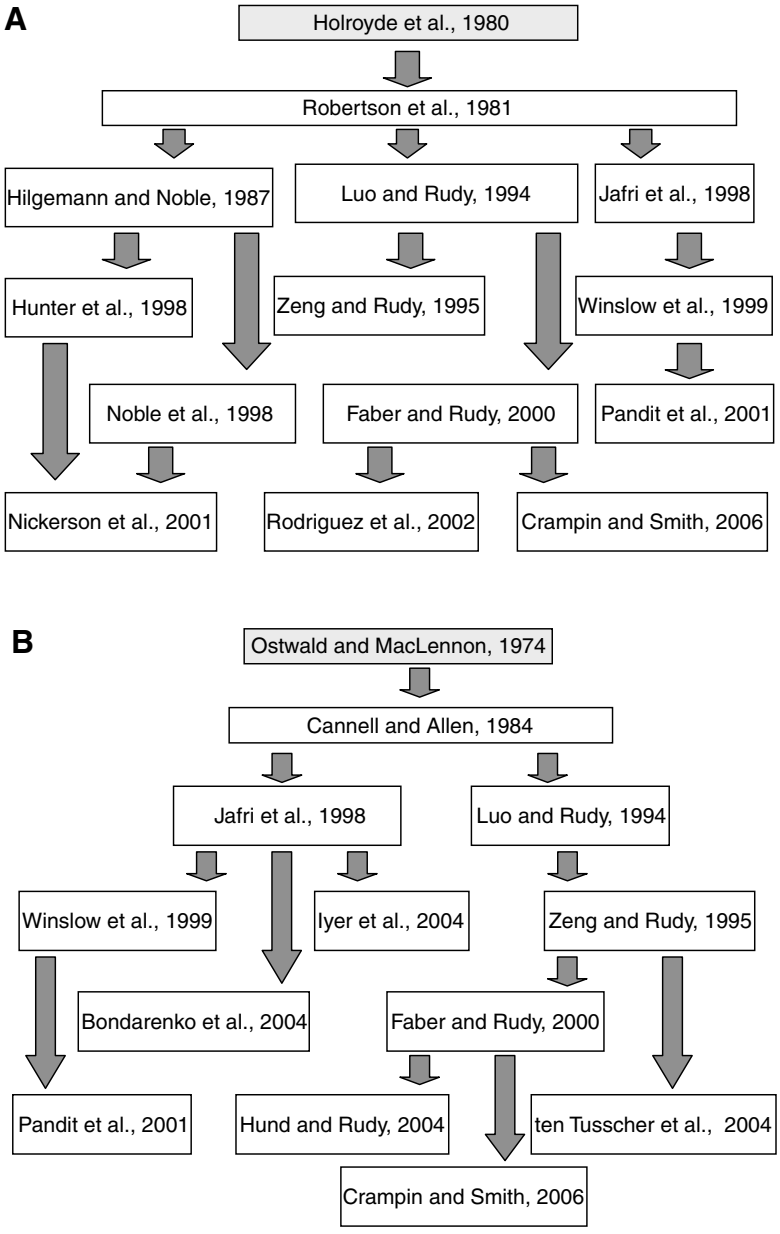

Fig. 4. Citation tree for the source of the (A) binding affinity of Ca^{2+} to troponin C and (B) binding affinity of Ca^{2+} to calsequestrin parameter, in cardiac myocyte mathematical models. Grey and white boxes indicate experimental and modelling studies, respectively.

To address issues 2–5 will require the community to build on these initiatives, and the development of openly available resources to disseminate models linked to the data sets used to parameterize them. We suggest that the following two types of entities should be collected and published online in a physiome database: published models, including complete codes for simulation, and peer-reviewed published data sets in accessible electronic formats. The first of these is the domain of the MIRIAM standard. Model entries in the database will be annotated using established ontologies, and include working and executable codes, using freely available tools, or

Table 1. *Review of computational models of cellular function*

Model/criteria	1	2	3	4	5	6	7
Smith and Crampin (Smith and Crampin, 2004)	NA	NA	NA	A	NA	A	A
Crampin and Smith (Crampin and Smith, 2006)	A	A	B	B	A	A	A
Niederer, Hunter and Smith (Niederer et al., 2006)	NA	NA	NA	B	NA	A	A
Wu et al. (Wu et al., 2007)	B	A	B	A	A	A	A

The models reviewed above classified according to the criteria outlined above, where the model satisfies (A), does not satisfy (B), or is not applicable (NA) when assessed against each criteria.

See text for details of each model.

computational code in an established language (C, Matab, Fortran, Pascal). These marked up executables with the addition of digitized data sets (see point 1 below) will ideally be available as part of the review process. This will enable the reviewer and user community to curate entries in the database with the following tools and criteria:

(1) Explicit links will be established between data sets and models. Specifically, each model will link to: (i) the data that were used to parameterize the model; (ii) additional data that are used to verify or demonstrate the scope and physiological application of the model; and (iii) known relevant data sets that the model does not satisfactorily fit. In addition each data set will link to: (i) model(s) that use the data set as part of the parameterization of those models; (ii) models that fit and/or help to explain the data set; and (iii) models that are not able to fit the data. These links will be edited by the authors.

(2) Classification of the model according to the objective criteria listed below. The authors will be invited to provide this classification. The ultimate goal is to have submission of a model to the Physiome resource with classification according to these criteria as part of the review process for major journals. Reviewers may be expected to verify the initial classification entered by the authors.

(3) A user feedback and review section where people can post non-anonymous 'amazon.com' style comments on their experiences. In each case the authors will be invited to provide a response and, if necessary, update their work.

Objective criteria

Below is the list of objective criteria that we propose for classification of computational models of cellular function. Each model is classified in each of the following categories as: (A) satisfies, (B) does not satisfy, or (NA) not applicable. This classification is not intended as a judgment on the validity of a given model or approach; but is intended to help define the scope and applicability of a model for potential users.

Objective characteristics of models

Biophysically based model criteria:

(1) Mass balance. The total mass of model variables leaving or entering the system is explicitly accounted for (and in the case of a closed system is conserved).

(2) Charge balance. Total charge of model variables leaving or entering the system is explicitly accounted for (and in the case of a closed system is conserved).

(3) Osmotic balance. Mass balanced model accounts for water fluxes and volume changes.

(4) Thermodynamic feasibility. Model components obey detailed balance and thermodynamic box constraints.

Criteria for comparing model to data:

(5) Initial conditions given for periodically driven models provide a beat-to-beat steady state (variables return to the initial condition after exactly one period).

(6) All parameter values are justified by cited experimental measurement, previous estimation based on model analysis, or based on model analysis in the current study, or qualitative commentary by the authors.

Computational documentation criteria:

(7) In addition to MIRIAM compliance, all model units are defined and used consistently and model initial and boundary conditions are unambiguously defined.

We now consider the models, from our own work, described above. The classification of each of the models against these criteria is given in Table 1.

Discussion

In the above section we have proposed a set of criteria for models in physiome databases, in addition to MIRIAM compliance, by which we hope to facilitate confidence in the use and reuse of biophysically based models of biological and physiological systems. These insist on a transparent connection between experimental data and model representation, and a set of objective model characteristics that will assist in quantifying the scope of a given model.

It would be naïve, however, not to consider the difficulties with implementing such a process. The culture of scientific publishing rewards the creation and publishing of new models rather than critiquing or reviewing existing work. The classification of models according to a set of criteria, as proposed above, may require significant investment of resources and, perhaps, requires new ways to recognize and to provide incentives for individual involvement.

As suggested in the MIRIAM proposal, an initial curation process will be most effective if performed by the model

author, rather than *post-hoc* by a separate curator. However, if models are to fulfill their role, giving qualitative (mechanisms) and quantitative (experimental data) understanding, it will be vital that there is a forum for an open and robust critique of models. This debate could take the form of challenging models with new data sets, as they become available, or critiquing modelling assumptions or approaches used in deriving a model. Developing a forum that encourages open debate amongst experts and users and provides useful information for non experts, while minimizing unproductive conflict, would clearly require skilled mediation and a well established code of conduct. However, as argued in the Introduction, we believe this type of curation will be an essential process for the ongoing development of integrated computational models

We have outlined a preliminary plan that expands the currently proposed criteria for model curation and we assessed four models from our own work against the proposed criteria. We hope that this proposal will itself generate dialogue and debate within the biological modelling community. Our five criteria for model assessment have been selected for their primary relevance to metabolic and electrophysiological models. However, any 'final' set of criteria must of course be selected and adopted by the community, and may possibly require the formulation of additional criteria, or even of alternative lists for the classification of models based on other frameworks, e.g. network inference models for gene–gene interactions, or signalling pathways. We see this goal as falling firmly under the aegis of the Physiome Project; motivated by the pressing need to establish standards to facilitate communication and debate about models, to accelerate the use, implementation and review of models and their connection with data by the scientific community.

The authors would like to thank Professor Peter Hunter, for helpful discussions. This work was supported. by the Marsden Fund of the Royal Society of New Zealand through grant No. 04-UOA-177 and National Institutes of Health grant No. EB005825.

References

Bassingthwaighte, J. B., Beard, D. A. and Li, Z. (2001). The mechanical and metabolic basis of myocardial blood flow heterogeneity. *Basic Res. Cardiol.* **96**, 582-594.

Bondarenko, V. E., Szigeti, G. P., Bett, G. C. L., Kim, S.-J. and Rasmusson, R. L. (2004). Computer model of action potential of mouse ventricular myocytes. *Am. J. Physiol.* **287**, H1378-H1403.

Buist, M. L., Cheng, L. K., Sanders, K. M. and Pullan, A. J. (2006). Multiscale modelling of human gastric electric activity: can the electrogastrogram detect functional electrical uncoupling? *Exp. Physiol.* **91**, 383-390.

Cannell, M. B. and Allen, D. G. (1984). Model of calcium movements during activation in the sarcomere of frog skeletal muscle. *Biophys. J.* **45**, 913-925.

Crampin, E. J. and Smith, N. P. (2006). A dynamic model of excitation-contraction coupling during acidosis in cardiac ventricular myocytes. *Biophys. J.* **90**, 3074-3090.

Crampin, E. J., Halstead, M., Hunter, P., Nielsen, P., Noble, D., Smith, N. and Tawhai, M. (2004). Computational physiology and the physiome project. *Exp. Physiol.* **89**, 1-26.

Cuellar, A. A., Lloyd, C. M., Nielsen, P. F., Halstead, M. D. B., Bullivant, D. P., Nickerson, D. P. and Hunter, P. (2003). An overview of CellML 1.1, a biological model description language. *Trans. Soc. Model. Simul. Int.* **79**, 740-747.

Faber, G. M. and Rudy, Y. (2000). Action potential and contractility changes in Na^+ (i) Overloaded cardiac myocytes: a simulation study. *Biophys. J.* **78**, 2392-2404.

Hilgemann, D. W. and Noble, D. (1987). Excitation-contraction coupling and extracellular calcium transients in rabbit atrium: reconstruction of basic cellular mechanisms. *Proc. R. Soc. Lond. B Biol. Sci.* **230**, 163-205.

Holroyde, M. J., Robertson, S. P., Johnson, J. D., Solaro, R. J. and Potter, J. D. (1980). The calcium and magnesium binding-sites on cardiac troponin and their role in the regulation of myofibrillar adenosine-triphosphatase. *J. Biol. Chem.* **255**, 1688-1693.

Hund, T. J. and Rudy, Y. (2004). Rate dependence and regulation of action potential and calcium transient in a canine cardiac ventricular cell model. *Circulation* **110**, 3168-3174.

Hunter, P. J. and Borg, T. K. (2003). Integration from proteins to organs: the Physiome Project. *Nat. Rev. Mol. Cell Biol.* **4**, 237-243.

Hunter, P. J., McCulloch, A. D. and ter Keurs, H. (1998). Modelling the mechanical properties of cardiac muscle. *Prog. Biophys. Mol. Biol.* **69**, 289-331.

Hunter, P., Smith, N., Fernandez, J. and Tawhai, M. (2005). Integration from proteins to organs: the IUPS Physiome Project. *Mech. Ageing Dev.* **126**, 187-192.

Iyer, V., Mazhari, R. and Winslow, R. L. (2004). A computational model of the human left-ventricular epicardial myocyte. *Biophys. J.* **87**, 1507-1525.

Jafri, M. S., Rice, J. J. and Winslow, R. L. (1998). Cardiac Ca^{2+} dynamics: the roles of ryanodine receptor adaptation and sarcoplasmic reticulum load. *Biophys. J.* **74**, 1149-1168.

Le Novere, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P. et al. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509-1515.

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

Luo, C. H. and Rudy, Y. (1994). A dynamic-model of the cardiac ventricular action-potential. 1. Simulations of ionic currents and concentration changes. *Circ. Res.* **74**, 1071-1096.

McCulloch, A., Bassingthwaighte, J., Hunter, P. and Noble, D. (1998). Computational biology of the heart: from structure to function. *Prog. Biophys. Mol. Biol.* **69**, 153-155.

Nash, M. P. and Hunter, P. J. (2000). Computational mechanics of the heart – from tissue structure to ventricular function. *J. Elast.* **61**, 113-141.

Nickerson, D. P., Smith, N. P. and Hunter, P. J. (2001). A model of cardiac cellular electromechanics. *Philos. Trans. R. Soc. Lond. A* **359**, 1159-1172.

Nickerson, D., Smith, N. and Hunter, P. (2005). New developments in a strongly coupled cardiac electromechanical model. *Europace* **7**, S118-S127.

Niederer, S. A., Hunter, P. J. and Smith, N. P. (2006). A quantitative analysis of cardiac myocyte relaxation: a simulation study. *Biophys. J.* **90**, 1697-1722.

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

Ostwald, T. J. and MacLennan, D. H. (1974). Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum. *J. Biol. Chem.* **249**, 974-979.

Pandit, S. V., Clark, R. B., Giles, W. R. and Demir, S. S. (2001). A mathematical model of action potential heterogeneity in adult rat left ventricular myocytes. *Biophys. J.* **81**, 3029-3051.

Pullan, A. J., Cheng, L. K., Nash, M. P., Bradley, C. P. and Paterson, D. J. (2001). Noninvasive electrical imaging of the heart: theory and model development. *Ann. Biomed. Eng.* **29**, 817-836.

Qian, H., Beard, D. A. and Liang, S. D. (2003). Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* **270**, 415-421.

Ribba, B., Tracqui, P., Boix, J. L., Boissel, J. P. and Thomas, S. R. (2006). QxDB: a generic database to support mathematical modelling in biology. *Philos. Trans. R. Soc. Lond. A* **364**, 1517-1532.

Robertson, S. P., Johnson, J. D. and Potter, J. D. (1981). The time-course of Ca^{2+} exchange with calmodulin, troponin, parvalbumin, and myosin in response to transient increases in Ca^{2+}. *Biophys. J.* **34**, 559-569.

Rodriguez, B., Ferrero, J. M., Jr and Trenor, B. (2002). Mechanistic investigation of extracellular K^+ accumulation during acute myocardial ischemia: a simulation study. *Am. J. Physiol.* **283**, H490-H500.

Schilstra, M. J., Li, L., Matthews, J., Finney, A., Hucka, M. and Le Novere, N. (2006). CellML2SBML: conversion of CellML into SBML. *Bioinformatics* **22**, 1018-1020.

Shaw, R. M. and Rudy, Y. (1997). Electrophysiologic effects of acute myocardial ischemia: a theoretical study of altered cell excitability and action potential duration. *Cardiovasc. Res.* **35**, 256-272.

Smith, N. P. and Crampin, E. J. (2004). Development of models of active ion transport for whole-cell modelling: cardiac sodium-potassium pump as a case study. *Prog. Biophys. Mol. Biol.* **85**, 387-405.

Smith, N. P., Pullan, A. J. and Hunter, P. J. (2000). Generation of an anatomically based geometric coronary model. *Ann. Biomed. Eng.* **28**, 14-25.

Smith, N. P., Buist, M. L. and Pullan, A. J. (2003). Altered T wave dynamics in a contracting cardiac model. *J. Cardiovasc. Electrophysiol.* **14**, S203-S209.

Smith, N. P., Nickerson, D., Crampin, E. J. and Hunter, P. J. (2004). Computational modelling of the heart. *Acta Numerica* **13**, 371-431.

ten Tusscher, K. H. W. J., Noble, D., Noble, P. J. and Panfilov, A. V. (2004). A model for human ventricular tissue. *Am. J. Physiol.* **286**, H1573-H1589.

Tomlinson, K. A., Pullan, A. J. and Hunter, P. J. (2002). A finite element method for an eikonal equation model of myocardial excitation wavefront propagration. *SIAM J. Appl. Math.* **63**, 324-350.

Trew, M. L., Caldwell, B. J., Sands, G. B., Hooks, D. A., Tai, D. C., Austin, T. M., LeGrice, I. J., Pullan, A. J. and Smaill, B. H. (2006). Cardiac electrophysiology and tissue structure: bridging the scale gap with a joint measurement and modelling paradigm. *Exp. Physiol.* **91**, 355-370.

Winslow, R. L., Rice, J., Jafri, S., Marban, E. and O'Rourke, B. (1999). Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure. II. Model studies. *Circ. Res.* **84**, 571-586.

Wu, F., Jeneson, J. A. L. and Beard, D. A. (2007). Oxidative ATP synthesis in skeletal muscle is controlled by substrate feedback. *Am. J. Physiol. Cell Physiol.* **292**, C115-C124.

Zeng, J. L., Laurita, K. R., Rosenbaum, D. S. and Rudy, Y. (1995). 2 components of the delayed rectifier K^+ current in ventricular myocytes of the guinea-pig type – theoretical formulation and their role in repolarization. *Circ. Res.* **77**, 140-152.

The Journal of Experimental Biology 210, 1492-1496
Published by The Company of Biologists 2007
doi:10.1242/jeb.02783

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at http://www.wikipedia.org/

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology.
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	chromatin immunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray.
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from chromatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as *cis*-acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of ‘C followed by G’ dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C–G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (*trans*-acting factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called ‘gene deserts’.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of gene and chromosome, the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the *Hox* genes in animals and *MADS-box* genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	insertion and deletion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon.
KEGG	The Kyoto Encyclopedia of Genes and Genomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. *Drosophila*, *Caenorhabditis elegans* and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3′ ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5′ and followed by a 3′ untranslated region (5′ UTR and 3′ UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA, siRNA and piRNA. Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA.

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA *in vitro*. The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3′ tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a ‘reflectron’ by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	RNA-induced silencing complex. A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC.
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR.
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters. Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
trans-acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called ‘mobile genetic elements’ or ‘jumping genes’.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3′ or 5′ flanking ends of the molecule (i.e. 3′ UTR and 5′ UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.