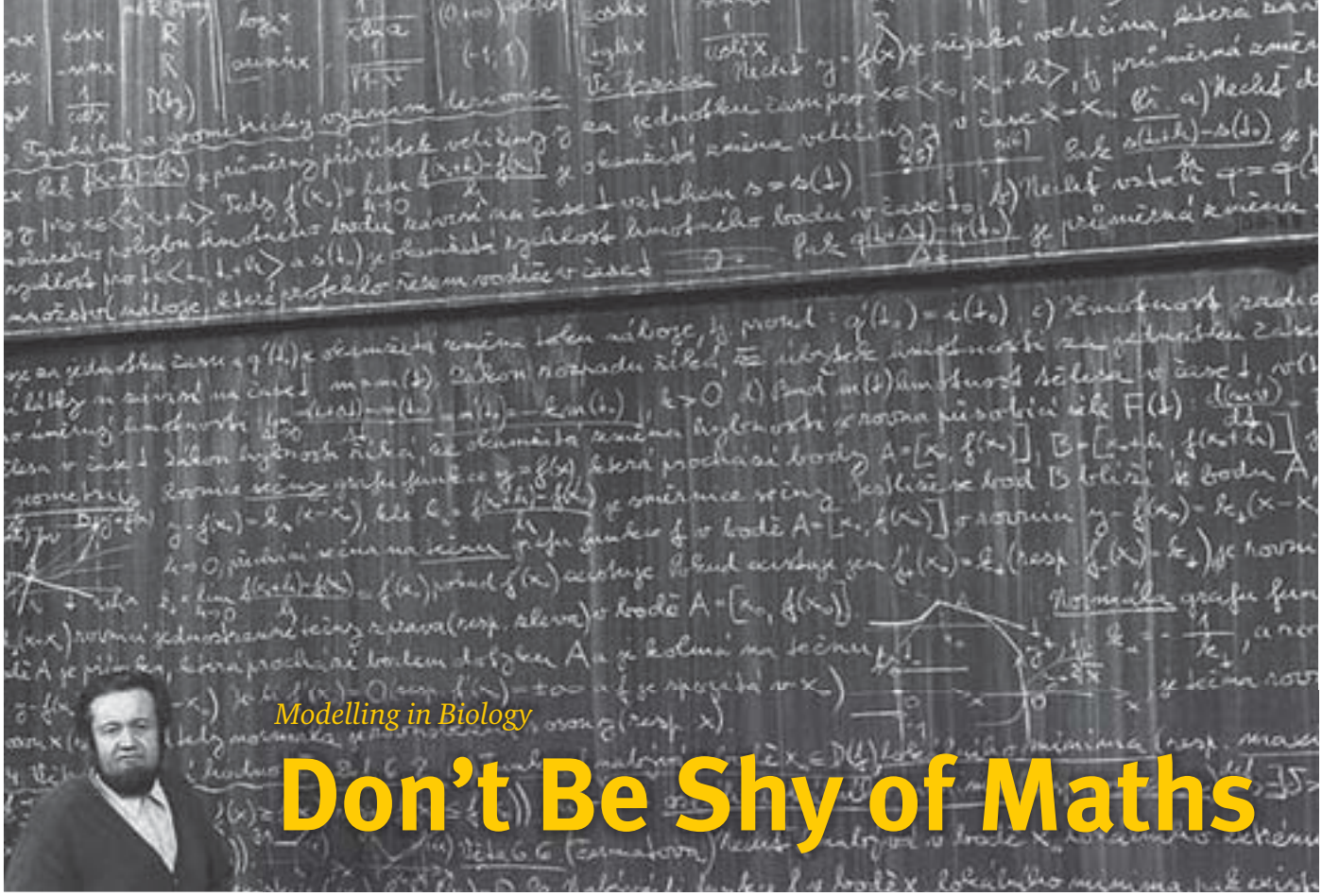


Modelling in Biology

Don't Be Shy of Maths

It's no longer possible to turn the explosively increasing datasets of modern research into biological insights without mathematical modelling. Fortunately, new software packages open this door to even the most entrenched mathophobe.

Biologists hate maths. Somehow we are more comfortable with our box-and-arrow pathways, with their intuitive feel of how causality flows from one molecule to the other, leading from cause to effect, stimulus to response.

However, times are rapidly changing. The explosive increase in the number and size of datasets generated by large-scale experiments, such as microarray, proteomics and metabolomics projects, is forcing a change in the way we model biological systems. Just how are we to turn these parts lists into insights into how cells, tissues and organs work? The human brain cannot cope with the complexity of the dynamic models needed to do this and so we are turning more and more to computer-operated mathematical models. Also it is not just large-scale pharmaceutical companies or dedicated math-biology labs that are doing this. New software packages are opening the door to the power of predictive modelling to even the most entrenched mathophobe.

Virtual man

Computer modelling does nothing more than scientists have always done down through the ages. They model systems using numbers and mathematical functions and look to see if the relationship between the inputs (stimuli) and the outputs (responses) match those of the real world. It is just that computers can handle much more complicated models than the poor human brain ever could.

Computerised biological models are growing, both in complexity and in predictive power. Massive models of metabolic and signalling networks are opening the way for rapid *in silico* design of new drugs. They can also test for their toxicity, at the same time reducing reliance upon the use of animal testing. A good model can save a lot of time and money.

Not without problems

Do they really work? Given the complexity of even a single cell and our ignorance of so much of what goes on in even the best-studied processes, can they really make useful predictions? Companies like Entelon think they can and they have the financial backing to prove it. Entelon is taking the power of predictive simulations seriously, developing huge bio-models of several human diseases. They harvest data from biomedical literature and assemble it into models called "virtual patients", a term which gives some idea of the confidence they place in the models' predictive power. Last month Entelon was granted a patent on a model of diabetes that is to be used to predict not only the system-wide effects of drugs, along with their optimal doses, but also to suggest the best paradigms for clinical trials.

Mathematical modelling in biology, however, is not without its problems. For a start, there is the issue of the quality of the data. Even the best model can be no better than the data it is built upon and this is particularly true for mechanistic models, in which every component has to be represented with some degree of accuracy. The problem is that there is a mismatch between the speed at which we can obtain certain data, such as protein expression levels or metabolite concentrations, and the time it takes to get other kinds of data, such as kinetic equilibrium constants. Plus it is often just these kinds of hard-to-get data that make all the difference to the predictive power of a mechanistic model.

Logical approaches

There are ways around this, however. Some labs have completely sidestepped this issue altogether, questioning even the need for such refinement of data. Last year, for instance, Julio Saez-Rodriguez and colleagues published a paper in *PLoS Computational Biology* in which they modelled the signalling network that controls activation of T-cells by T-cell receptors (*PLoS Comput. Biol. 3(8): e163*). Their model incorporated 94 nodes and 123 interactions and it accurately mimicked several of the published observations on this network. They then went on to test the predictive power of the model and found that it anticipated the con-

sequences of altering CD28 or knockout of FYN, which were subsequently validated experimentally. The interesting thing about their model, however, was that they did not model the activation levels of each node; they simply modelled them as being "on" or "off". The so-called logical approach. So it seems that in some cases at least, modelled systems can be so robust that they are immune to the specific details of their individual components.

Not just for Big Pharma

The limitation on the quality of data, along with our incomplete knowledge of the elements of the pathways, has spawned other alternatives to mechanistic models of cell signalling. Genstruct is a company that uses models in drug discovery but of a different kind. Instead of modelling every element in a signalling pathway, hoping that the elements they left out don't make a difference, Genstruct make use of an artificial intelligence approach which treats the details of the pathways simply as "black boxes". They start with a knowledge database built on experimental observations and pass this through an inference-making process to generate their model. This model in turn generates hypotheses about how a system should behave when perturbed in a specific way and the results of experiments testing these hypotheses are added back to the database. Does it work? One of the world's biggest pharmaceutical companies thinks it just might. This February Genstruct announced that they have signed a collaborative agreement with DuPont covering future efforts to predict the preclinical safety of drugs and their possible toxic effects on the liver.

However, modelling is not just for big industrial labs. Even very simple computer models can be a powerful tool for any biologist. In essence, models are a stringent test of whether your intuitive understanding of a process really makes sense. They force you to be specific and categorical about what you are saying.

There are many modelling software packages available, some of them free. They can take the mathematical pain out of modelling, usually requiring only the kind of knowledge you would find in any lab, such as an understanding of the basics of Michaelis-Menten kinetics or equilibrium. Some packages, such as *E-Cell*, have graphical modules, so you can build your model with a mouse, adding in the parameters as you go. The more serious modeller might invest in the popular *MATLAB* package, to which has recently been added the *SimBiology* module, specifically designed with the cell modeller in mind.

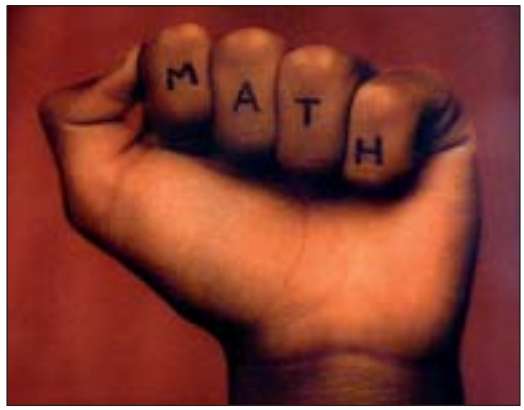
The link to experiments

Before you start doing your own modelling, however, there are a few questions you need to ask yourself. First of all, what is the scope of your model going to be? That is, how much of a pathway does your model need to take into account and which components can your model safely ignore without it affecting your predictions? If your model is going to be linked to carefully controlled experiments, then this question is largely answered by the particular details of your experimental design.

Once you have established the scope of your model, the next question is the level of detail. No-one can predict the properties of

a protein from its atomic structure, for instance, but packages like *M-Cell* do allow the diffusion of single molecules within a cell to be modelled. Most models can safely trade off the intense computational demand of such approaches by allowing a certain level of granularity in the model. Biological models often assume that, to some extent, processes are modular. Separate pathways generally behave in the same way, irrespective of what other pathways are doing, except where those other modules become inputs to the pathway itself.

The spatial aspect of cell function must also be encapsulated. Most models accomplish this by adopting a compartmental approach. They simulate the spatial dimension by allocating species (proteins, ligands, metabolites) into a finite set of compartments, giving them freedom to move from one compartment to another according to sets of rules, such as diffusion rates. Packages like *E-Cell* make defining these compartments almost as simple as drawing a circle with the mouse.



Robust and stable?

Will your model be deterministic or stochastic? A deterministic model limits the behaviour of molecules with strict rules, so that every time you run the model with the same starting conditions you get exactly the same results. This is of course less realistic but is computationally less demanding. On the other hand, stochastic models are built upon statistical rules and provide a more realistic simulation (which may be critical when the number of molecules in a compartment is low). The most sophisticated modelling software will even decide which components can safely be modelled deterministically and which must be modelled stochastically.

Once you have built your model, it must be tested for robustness and stability. In other words, even though your model might accurately predict a set of experimental observations, you want to be convinced that it will still do so even if you change some of the parameters. Experimental measurements are subject to several sources of error. How much will these affect the predictions of your model? In addition to checking for the stability of the model, checking for how it depends upon variations in a particular parameter can be of interest in its own right. Finding that a model is especially sensitive, say, to the concentration of a particular molecule or phosphorylation site can suggest a novel point of control in a biological process and so yield a testable hypothesis.

Pressure to publish

Even if you don't feel up to the challenge of making your own model from scratch, you can still benefit from other people's efforts. There is increasing pressure to publish biological models in repositories, in the same way as sequence data, which must be deposited as a condition of publication. This has been routine with molecular models for some years and is only now being encouraged with network models as well. The BiomodelDB hosts over 100 models and the CellML repository even more. These models can be downloaded or even run over the web, offering a greater level of scientific accountability since peers can test the robustness and repeatability of published data, as well as adapt them to their own research interests. In future, these models will be integrated into other databases, just as today protein details are linked to citations in PubMed.

Robot Scientist

Whatever the details of the model, models are essentially all used in the same way. There is a cycle of model refinement and hypothesis generation. Data are used to sculpt the model. The model predicts outcomes. These predictions are compared with experimental data and the difference used to further refine the model.

However, if computers can host the model, why not go all the way and computerise the whole experiment-model cycle? A team at Aberystwyth University has done just this with the Robot Scientist project (*www.aber.ac.uk/compsci/Research/bio/robotsci/intro/*). While there is nothing particularly new in having a robot run microplate experiments and analyse the data, the lab robot in the Robot Scientist project actually designs the experiments in the first place, selecting the protocol expected to yield the maximum information based on previous knowledge and findings. Does it work? In one test, Robot Scientist performed as well as human scientists and at greatly reduced cost (*www.tfot.info/column/1006/the-other-meaning-of-computer-science.html*).

The ways in which computer models are used are as diverse as the systems that they model. Academic biologists look to mechanistic models as a way of understanding complex biological systems and even of bridging the divide between biochemistry and physiology. Industry wants faster, cheaper and more reliable ways of predicting what effects a compound will have on complex and variable systems. However, whether computer modelling will change the way in which we describe and understand biological systems remains to be seen. After all, the most accurate model of a cell is, well, a cell . . .

STEVEN D. BUCKINGHAM

Websites for biological modelling

- **MATLAB SimBiology module** (*www.mathworks.com/products/simbiology/*)
  Comprehensive suite for modelling along with useful tutorials
- **BiomodelDB** (*www.ebi.ac.uk/biomodels/*)
  Repository of models with links to citations
- **CellML** (*www.cellml.org/*)
  XML-based modelling language and repository with tutorials
- **Robot Scientist** *www.aber.ac.uk/compsci/Research/bio/robotsci/*)
  The whole model/experiment cycle deployed by machine
- **Neuron** (*www.neuron.yale.edu/neuron/*)
  Package for simulating electrophysiology of neurones
- **Genesis** (*www.genesis-sim.org/GENESIS/*)
  Neuronal simulation package
- **Systems biology portal** (*http://systems-biology.org/resources/model-repositories/*)
  Systems-biology.org collection of model repositories and software collections