GC1 In Vivo–In Silico (iViS): the Virtual Worm, Weed and Bug

Ronan Sleep

We routinely use massively powerful computer simulations and visualisations to design aeroplanes, build bridges and to predict weather. With computer power and biological knowledge increasing daily, perhaps we can apply advanced computer simulation techniques to realise computer embodiments of living systems. This is the futuristic proposition of a research challenge proposed by UK computer scientists. The project, called *in Vivo – in Silico* (iViS) aims to realise fully detailed, accurate and predictive computer embodiments of plants, animals and unicellular organisms.

Initially the aims will be restricted to simple and much studied life-forms such as the nematode worm, the humble weed Arabidopsis, and single cell organisms such as streptomyces and bakers yeast: hence the subtitle 'the virtual worm, weed and bug'. These model organisms, apparently so different, have much in common:

> *As we trace the increase of complexity from single cell creatures, through small animals like worms and flies . . . evolution is not so much adding new genes performing wholly new functions - what it's chiefly doing is to increase the variety and subtlety of genes that control other genes*[1]

Further, the human and worm have a common ancestor from which we jointly inherit many similar genes. An example is the recent discovery that there is a gene in the worm that is similar to the human breast and ovarian cancer gene BRCA1 (Boulton et al, Current Biology, V14, No.1 pp33-39). So there is considerable hope that studies of the simpler life forms will have real relevance to humans.

Possible benefits of iViS include an understanding of regeneration processes in plants and animals, with potentially dramatic implications for disease and accident victims. But iViS may also lead to revolutionary ways of realising complex systems: instead of designing and programming such systems in excruciating detail, perhaps we can just *grow* them from compact initial descriptions in a suitable medium. We know it's possible, because that's exactly what nature does with the worm, the weed and the bug.

The vision

iViS offers a powerful vision in which future life scientists can take virtual reality fly-through tours of a plant, animal or colony of cells, studying what is happening at scales ranging from whole life-form to what goes on inside an individual cell, and stretching or shrinking time. Filters control what is seen by the observer, allowing concentration on specific aspects such as cell division, motility or chemical potential.

This is an attractive way of exploring our knowledge about a life-form. But iViS may offer more: with sufficient effort, it might be possible to raise the faithfulness of the underlying model to the point where it becomes predictive as well as descriptive.

[1] `http://www.sanger.ac.uk/HGP/publication2001/facts.shtml`

If this happens, it will become possible to perform meaningful observations and experiments in Silico. And we want to cover a wide range of phenomena: specifically, we include: development from an initial fertilized cell to a full adult, cell function and interaction, motility and sensory behaviour, including interactions with other life-forms. Virtual experiments (e.g. moving a virtual cell during development) should lead to the same outcomes as real life.

iViS and the life science data mountain

Computers are vital to the Life Sciences: they record, process, visualise and automatically distribute data, and even design and run experiments. They analyze the results in terms of large statistical and other mathematical models of biological processes.

But what do all the numbers, graphs, and sphaghetti-like diagrams that emerge from the latest experiments all mean, and what can we do with this data? Making it all fit together into a coherent and useful picture presents a major challenge - many biologists would say *the* major challenge - facing the life sciences.

This problem is now so pressing that the UK's Biotechnology and Biological Science Research Council (BBSRC) is establishing a number of Centres for Integrative Systems Biology. These Centres will need the vision, breadth of intellectual leadership and research resources to integrate traditionally separate disciplines in a programme of international quality research in quantitative and predictive systems biology. iViS offers a challenging focus of attention for such centres.

iViS as a driver for global knowledge integration

Part of the answer to the data mountain may lie in the way in which the world wide web is revolutionising our approach to knowledge organisation. The web is already a vital window on the world for scientists wishing to remain up to date. Groups of laboratories that previously worked at arms length and communicated infrequently only via journals and the conference circuit now converse via the web within seconds, swapping massive datasets to compare results. Scientists have begun to exploit the web in its own right by establishing global Virtual Knowledge Repositories to share data, theories and models.

A particularly relevant example is Physiome[2], which supports

> *the databasing of physiological, pharmacological, and pathological information on humans and other organisms and integration through computational modelling. 'Models' include everything from diagrammatic schema, suggesting relationships among elements composing a system, to fully quantitative, computational models describing the behaviour of physiological systems and an organism's response to environmental change.*

Virtual Knowledge Repositories like Physiome will help reduce the proliferation of models and theories that explain parts of the global mountain of life science data. But this in turn will create a deeper challenge: instead of fitting raw data pieces together, we will be faced with the problem of making the models fit into a consistent

[2] `http://www.physiome.org/`

larger model. Sometimes this will be easy, for example when there is a simple input-output relationship between subsystems. More often —perhaps the rule rather than the exception— combining two models will show unexpected interactions inconsistent with in vivo data.

Part of the problem is that mathematical models deal in numbers, devoid of meaning. The latest evolution of web technology — the semantic web — may help fix this. There is now provision for the web to enhance raw data with additional information called metadata. This can tell the recipient what the data means, how it is represented, the way in which it was generated. Models, which often come in the form of a computer program, can be tagged with metadata describing their assumptions and use: effectively an inbuilt instruction manual.

There are already over 40 metadata dictionaries[3](called ontologies) for the life-sciences. So the drive and energy to create bio-ontologies is already very active. But there is not the same drive to draw them together into a unified whole. The iViS challenge provides just such drive, because the in Silico modelling of a *complete* life-form, will require harmonious working across all relevant ontology boundaries.

Even if we can build a simulation of a life-form that successfully integrates all known data, we need to take care in choosing our models. If they follow all the raw data too closely, the models may lack any predictive power. For example, we can always find a polynomial of degree $(n-1)$ to fit n data points exactly. This is a strong reason for complementing data-driven modelling work on iViS with more abstract top-down approaches. If we take care there will be at least some domains which succumb to iViS's whole life form modelling approach: developmental biology looks a good bet.

Meeting the Challenge: iViS research themes

The obvious targets for iViS models are the organisms selected for special attention by biologists for over a century. These range from single cell life-forms such as yeast or streptomyces, through model plants such as Arabidopsis and maize to creatures such as the nematode worm, the fruitfly, and the squid.

But how can we 'breathe life into data' via computer simulation? This is not simply a question of computing speed or memory, but how to represent the mass of known data as a set of interacting computational processes. We can get computers to simulate massively complex aircraft or bridges, but getting them to grow a worm, weed or bug is significantly beyond the current state of the art.

Nevertheless, it may not be impossible if we build determinedly on the considerable body of work underway to explore ways of organising life science data. One example is the Edinburgh Mouse Atlas Project[4]:

> *The EMAP Atlas is a digital Atlas of mouse embryonic development. It is based on the definitive books of mouse embryonic development ... yet extends these studies by creating a series of interactive three-dimensional computer models of mouse embryos at successive stages of development*

[3] `http://obo.sourceforge.net/`
[4] `http://genex.hgu.mrc.ac.uk/`

with defined anatomical domains linked to a stage-by-stage ontology of anatomical names.

It can be expected that growing numbers of life science virtual knowledge centres will follow EMAP in adopting some form of spatio-temporal framework. The role of iViS is to expand this vision to a dynamic 3-D working model, initially targeting much simpler life-forms.

There are a number of research strands in the Computing Sciences needed to support the aspirations of iViS. We might bundle them under the heading: *Computational Models and Scaleable Architectures for in Silico Life Sciences*. Some strands will work bottom-up, paying great attention to biological data. Other strands will work top-down, studying minimal abstractions capable of generating the phenomena exhibited in vivo. Many will work 'middle-out', balancing the desire to be simple, elegant and general with the desire to be faithful to the data.

Key to success will be the development of a new breed of computer languages for representing and manipulating biological data in a meaningful way, and using it to drive a realistic, highly detailed, simulation which can be explored using advanced interfaces.

Groups of computer scientists are already exploring new languages, architectures, and system design and analysis tools for the life sciences. Luca Cardelli of Microsoft[5] gives a good picture of this work. Cardelli and others are tackling the complexities of life science head on, developing industrial-quality models aimed at handling the masses of detail in a living system, and - of critical importance if the results of iViS are to be trusted - validating the resulting models.

The impact of such work on the life sciences could be as dramatic as the discovery of structured programming was for computing in the late 1960's. Prior to structured programming, even short programs looked like a mass of spaghetti, just as much of our knowledge of living systems does now. If biological analogues of the compositional primitives of structured programming (sequencing, alternation, and repetition) could be discovered, the prospects for integrated systems biology would be very bright indeed.

Such direct attacks on the compexity of biological detail are complemented by more abstract top-down approaches. These begin by asking what sort of computational systems have emergent life-like properties. Such abstract models can be useful when viewing some particular aspect of a plant, animal or cell. For example Prusinkiewicz[6] has almost created a new art form for constructing good-looking pictures of plant growth from remarkably simple abstract models called L-systems. These models capture only a tiny part of the truth, but iViS may need such simplifying ideas to help structure the great mass of detail.

What about raw computer power and advances in haptic and other interface technologies? Certainly they will be needed: but some will emerge anyway from the computer industry without special prompting. The critical problem is to get the underlying computational models right: once we have these, we can begin —as it were— serious work on designing the building and sorting out exactly which of the contemporary technologies we should use to actually construct an iViS prototype.

[5] `http://www.luca.demon.co.uk/`

[6] `http://www.cpsc.ucalgary.ca/Research/bmv`

Demonstrators and outline roadmap

iViS progress will surface as a number of key demonstrators, culminating in whole life form models covering a wide range of phenomena. Intermediate demonstrators will cover a narrower range. Modelling the development of form during development is one example, motivated by the following quote:

> *Perhaps no area of embryology is so poorly understood, yet so fascinating, as how the embryo develops form. Certainly the efforts in understanding gene regulation have occupied embryologists, and it has always been an assumption that once we understand what building blocks are made, we will be able to attack the question of how they are used. Mutations and gene manipulations have given insight into what components are employed for morphogenesis, but surely this is one example where we need to use dynamic imaging to assess how cells behave, and what components are interacting to drive cell movements and shape changes*(Scott E. Fraser and Richard M. Harland, Cell, Vol. 100, 4155, January 7, 2000)

A speculative timeline is:

> *within 5 years:* early results on developmental phenomena in plants and animals, and first unicellular demonstrations. Prototype modelling frameworks and validation methodologies.
>
> *within 10 years:* first prediction of a textbook result from an assembly of component models; models of meristem growth; models of simple animal development; reliable unicelular models; mature iViS modelling environments.
>
> *2017, 100 years after the publication of D'Arcy Thompson's paper 'On Growth and Form'* first substantial demonstration of iViS whole life model.
>
> *within 20 years:* iViS models in common use.

What happens next?

A website[7]for the iViS Grand Challenge has been established, and funding is being sought to help build the communitity of cross disciplinary scientists needed to elaborate and vigorously address this very exciting challenge.

[7] `http://www.cmp.uea.ac.uk/ivis`