Can biological phenomena be understood by humans?

Despite the increasingly successful collaboration between physics and biology, there are contexts in which their styles and philosophies can diverge. Nowhere more so than in 'understanding'.

"It is nice to know that the computer understands the problem. But I would like to understand it, too." So said the theoretical physicist Eugene Wigner when shown the results of a large quantum mechanics calculation. Now, if a theoretical physicist of a younger generation is to be believed, today's biologists are in danger of delegating to their computers the job of understanding biology. Physics Nobel laureate Robert Laughlin — who was able to understand the fractional quantum Hall effect without the aid of computers — was invited to a conference in San Diego earlier this month on quantitative challenges in the post-genome-sequence era. His job was to stir things up, and he duly did so. His remarks sparked a spirited defence of the post-genome-sequence biological agenda, but also prompted some useful contemplation of just what understanding means — or may come to mean — in biology.

To understand Laughlin's complaint, take the field of structural genomics. Here the agenda is to work out the structures of all the proteins encoded by an organism's genes, on the premise that a protein's structure can be a useful guide to its function. The problem is that protein structures are hard to obtain: the number of experimentally determined structures will never catch up with the number of known sequences. But there turns out to be great redundancy in the structures adopted by protein domains: only a limited number of structural arrangements (or 'fold' types) are found in nature. So the plan is to obtain the structures of all the fold types and then to model proteins of unknown structure into these folds on the basis of similarity of sequence, chemical plausibility and so on. In this way, it should be possible to model most globular protein structures with useful accuracy by the time the Human Genome Project has been completed in 2003.

Dangers of the black box

So far, so good, and this seems an eminently sensible way to identify new candidate drug targets, or to generate hypotheses about protein function that can then be tested by experiment. But if structural biologists then have a computational black box into which they can feed an amino-acid sequence, and get a protein structure out of the other end, can they say that they understand protein folding?

Laughlin would say no. He looks at a protein and sees an "emergent phenomenon", an entity whose properties cannot be derived from the properties of its parts. So, just as the behaviour of a convecting fluid cannot be predicted from analysing its component molecules, so the function of a protein does not reside in the properties of its amino acids. Moreover, he argues that the higher-order organization (the protein's structure and function) is insensitive to the nature of the parts — the details don't matter. In one respect this is certainly true of proteins, in that totally different sequences of amino acids can yield the same fold structure, and thus the same function. But proteins can also be exquisitely sensitive to some subset of the details: change one amino acid (if it's the right one), and the protein function can be destroyed. The challenge is to find out which details count, and why.

Laughlin worries that in the post-genome era, biologists will be too busy accumulating facts and modelling them to seek simplification and underlying principles. The fact-gathering tendency is apparent not just in structural genomics, but also in functional genomics (where the aim is to identify the role of each gene in the genome) and proteomics (with a similar aim for each protein in the cell or organism). Evidently, there are enough facts to keep biologists busy gathering them for decades, so when will they have time to think?

In defence of modelling

One answer, conveyed forcefully at the conference by Klaus Schulten (for protein-structure studies) and David Botstein (for genomics), is that modelling aids thinking, by helping to generate hypotheses. Where a theoretical physicist might gain intuition from pencil and paper, a structural biologist may need molecular dynamics simulations. (And, to be fair, physicists who work on complex systems are themselves no strangers to computer modelling.) The microarrays that are becoming the workhorses of genomics can produce a million data points in a single study; even a physicist, argued Botstein, would find it hard to make sense of these data as a table of numbers. So it makes sense to use the computer to put the data in some kind of order before trying to think about them. The ordering may be driven by a hypothesis — for example, one might look for evidence of a periodic cell cycle in the transcription of genes — but in the absence of a hypothesis, the data themselves may suggest a pattern.

Although one structural biologist went so far as to proclaim that "we have to free ourselves from the hypothesis-driven approach", even he admitted that it would be nice to find some underlying simplicity. But he and others expressed the view that a search for unifying principles would be premature — that it will first be necessary for biology to go through a stage of fact accumulation and pattern recognition.

But what if it's not just a 'stage'? Is it possible that the parts will be enumerated and the functions found, and still there will be no simplification? Fortunately, enlightenment can come in different forms — not just in the elegant simplicity of a physicist's theory, but also in the more utilitarian guise of an engineer's analysis. As Hartwell *et al.* have argued in the recent supplement *Impacts of Foreseeable Science* (*Nature* **402,** suppl. C47–C52; 1999), molecular and cell biology may have more in common with engineering and computer science than with the basic sciences; for example, the kind of modelling needed to understand the complex intracellular networks that underlie most biological functions comes straight from engineering control theory.

As shown by two papers in last week's issue (see *Nature* **403,** 335–338, 339–342; 2000), it is becoming possible not just to analyse naturally occurring networks in this spirit, but also to design and build biological networks to implement desired functions. That, surely, is a kind of understanding worth having, and one that theoretical physicists can recognize as progress of a sort. ■