

Towards more measurement in biology

Even the best known systems in biology may not be as well-understood as is generally believed, which means that understanding is incomplete and may even be misplaced.

WHAT does one do when challenged to justify in public what *Nature*, or at least this small part of it, has been preaching about the need for more quantitative observation and analysis in molecular biology? What better preparation could there be than to reread Mark Ptashne's "*A Genetic Switch*" (Cell Press and Blackwell; 1986)? For there is no more vivid account of a deliciously neat device in molecular biology than this tale of how bacteriophage λ infecting an *E. coli* cell can either kill its host by multiplying at its expense or can be incorporated in the host genome, cryptic until an insult is delivered to the cell.

Nothing in the plea for a greater regard for quantitative data is meant to diminish molecular biology. On the contrary, as Ptashne's book shows clearly, the structure of DNA provided not just the answer to the question "What is the physical embodiment of inheritance", but solved the problem of the physical mechanism linking genotype and phenotype. Schrödinger's "*What is Life*", rich though it was in speculation about the nature of genes, said little about the need that genes should determine the physical characteristics of the cells in which they are found. And even Watson and Crick, in their celebrated throw-away line, "It has not escaped our attention...", failed to remark that they had revolutionized biochemistry.

Yet molecular biology remains a largely descriptive science. In the various human genome projects, for example, the objective is essentially to list the genes and to specify their nucleotide sequences, as well as the sequences of the regions of DNA that hold the genes together in the chromosome. The naming of parts does not by itself yield understanding.

None of that implies that the human genome projects will tell us nothing about the working of the cell. On the contrary, there is every likelihood that clever people poring over the listings as they accumulate will learn, simply by inspection, a great deal about the life of the cell. The mere arrangement of regulatory genes must say something important about gene regulation. And it will be invaluable to learn, by the comparison of different genomes, how genetic rearrangement contributes to speciation. And there will be practical benefits, some gene therapy, for example.

So the complaint that molecular biology is insufficiently quantitative at present is a modest and subtle complaint: molecular biology would be in even better condition than it were otherwise. The lysogeny of *E.*

coli is a good illustration because it is such an arresting tale that has been so well described by Ptashne.

The essence of the tale is easily told. Of λ 's 17 genes, one produces a protein called repressor whose effect, when appropriately bound to viral DNA, is to ensure that only repressor remains active. That is the case whether the λ gene is incorporated in the host genome as a linear double-stranded stretch of DNA as in lysogeny or, alternatively, is formed into a circle as when injected into a cell from an external virus particle. Either way, no other λ genes than repressor can function.

The elegance of the mechanism rests on a second λ gene called *Cro* which, if there is no repressor protein present to repress it, will bind to the same DNA sites and repress the activity of repressor. This *ying-yang* way of doing business is generally taken as a paradigm for the working of switching mechanisms in other gene systems. How well is it explained?

Qualitatively, there are no problems. In lysogenized cells, there is repressor protein to switch off all but repressor in infecting λ DNA, so that viral replication and destruction of the host cell by lysis do not follow. The difficulties are quantitative. But first one needs to know that the functional form of the repressor protein is a dimer of the native molecule, and that repressor dimers can bind to three contiguous sites on λ , arranged in order of decreasing affinity. When bound to the most attractive site, the protein increases the affinity of the second site and their twin occupancy prevents transcription of all viral genes except repressor itself. But if the third site, with the lowest affinity, is occupied, repressor production ceases. It is reckoned that the saturation of all three sites occurs when there are 100 or so free molecules in the cell.

What would a physical chemist make of this? The first observation would be that there is a cascade of chemical equilibria: that between the monomers and the dimers, then between cytoplasmic dimers and the three successive stages of occupancy of the DNA binding sites. On the face of things, the law of mass action applies. The concentration of dimers will be proportional to the square of the concentration of monomers, and similar arguments will apply to the occupancy of the three genomic sites. But is that likely to apply when there are only 100 or so molecules?

That is where the difficulties arise. Even to assess whether mass action applies re-

quires consideration of such things as the diffusion rates of repressor molecules in the cytosol and probably, in the last resort, a model for the distribution of ribosomes (where protein molecules are synthesized) relative to the three active sites. But so far as can be told, these assessments have not been made, nor has there been a serious attempt to calculate the likely fluctuations of the performance of 100 molecules of repressor protein — which may be vital for an understanding of the function of the genetic switch (not to mention the likelihood that the cell concerned will escape lysis).

Indeed, the chances are that this delicate genetic switch is not a phenomenon of equilibrium thermodynamics at all, but a kinetic one. And if that is possible in this rudimentary case, how much more likely is it that, in the functioning of the transcription factors that transcribe genes in eukaryotic cells, kinetic considerations will be dominant? There, the active component is a polymerase molecule, but there are typically at least six others involved, each of which is added to the complex in an apparently precise order. How can that be arranged? Or, to turn the argument around, if the law of mass action applied, the required concentration of the last of these protein components would probably be orders of magnitude greater than that of the first, which is not the case.

This apparently arcane argument is not irrelevant. There is at least one well-known case, that of the chemical composition of interstellar gas, where equilibrium arguments give answers radically different from the measurement. The point is important not merely because we need to know what is really happening, but because, if the behaviour of repressor is determined kinetically and not by equilibrium thermodynamics, all the energetic implications of what appear to be equilibrium constants may be spurious.

But this is not the only source of uneasiness about the state of what must be one of the most elegant of all the tales molecular biology has to tell. There remains the problem of knowing how the molecules of repressor fold into their characteristic dumb-bell shape, why dimers are so much more stable than monomers and how an α helix in the amino-component interacts with DNA at the binding sites. To be more precise, *what* happens has been determined by elegant genetic experiment; *why* that, not something else, happens remains to be discovered.

John Maddox

*Gist of a talk at the American Institute of Medical and Biological Engineering, Washington DC, 7 March 1994.