The Preoccupations of Twenty-First-Century Biology

*David Baltimore**

The life sciences are in transition, with the beginning of the 21st century symbolically marking the initiation of this redefinition. It is a creative, exciting repositioning because it derives from a changed perspective on what is possible. The appearance of the human genome sequence at the turn of the millennium ushered in the post-genomic era, an era of new possibilities, new challenges and new paradigms.

Until the millennium, biologists struggled to produce data. Each of us had his or her particular perspective and each toiled to flesh it out by laboriously finding relevant proteins and genes, learning how they worked and trying to link them to the larger events of biological systems. Now we are reveling in a data and materials glut. All the genes and thus all the proteins are catalogued. Gene expression profiles in particular cell types are available online and comparisons of normal and diseased tissues are appearing daily. The sequenced genome in bite size pieces is readily accessed and an increasing fraction of the genome has been knocked out in available mice. And oh, the wonderful new machines we have to work with!

Consequently, life science research has become both easier and more difficult. Easier because we no longer have to clone a gene to get access to it; we can buy it from a supplier. If one wants to know something about a

*Division of Biology, California Institute of Technology, 1200 E. California Blvd., MC 147-75, Pasadena, CA 91125, USA, e-mail: baltimo@caltech.edu.

property of a protein, such as its structure, or one of a related protein, one can download it and study its molecular details. But the field is egalitarian: a scientist in India or California or London has the same access to data and materials. This makes originality harder to come by. Previously, one might have had a niche in biology to oneself, maintaining the hegemony by running ahead of the pack with a big laboratory of hard workers. But today such a strategy does not work because the access to tools and materials allows one to leap frog into the fray. Thus, there is a higher premium on innovative thinking, on the creative asking of questions.

The edge today goes to people who see a new avenue of application of the ubiquitous tool bag. One way to find open avenues is to focus on recently discovered molecules whose functions are up for grabs. The post-docs in my laboratory have increasingly been seeing such an avenue in microRNAs. Only found a few years ago and still in the process of being catalogued, the biological roles of these extraordinarily powerful little strands of RNA have yet to be extensively investigated, providing the post-docs with an open field. Actually, the field is only open temporarily; it is rapidly becoming crowded as more and more investigators ask how their pet area of study might be impacted by these previously unrecognized cellular components.

To study even a single microRNA, one uses all the tools of the moment: the array technologies, the data libraries, the sequence compendia. Mountains of data can be the investigator's friend when he or she approaches them from a well-defined point of view. But the data have stories to tell of their own. These stories are not about single entities but about relationships, correlations, interactions, control mechanisms, all of which can be teased from the data themselves. In the best of circumstances, the analysis leads to a hypothesis that can be tested by other data: data that may have to be collected. In this way, the data glut gets mined for significance and pathways of knowledge are unveiled. But more importantly, the new knowledge may suggest new perturbations that can be introduced into biological systems and then interrogated for their consequences. This back and forth between data, hypotheses, experimental design, collection of new data and then verification or falsification of hypotheses is a rich and satisfying process, making contemporary biological science particularly powerful.

But this data mining and hypothesis construction is only a part of the 21st century scene. There is structural and mechanistic biochemistry, which benefits so greatly from the increased computing power we all enjoy. Protein folding, that tantalizing mystery at the core of biology, is also yielding to the power of digitalization and computation. Structural biology promises to finally deliver in this century its long-awaited fruits of precise knowledge of the molecular events that underlie the remarkable ability of linear combinatorial association of 20 amino acids to make the universe of nanomachines that carry out the processes of the biological world. And new and very expensive tools will give us insight into the aggregates of proteins that carry out the events of biology: the proteins of the synapse that send signals among neurons, the proteins of transcriptional control that give individuality to cells, the proteins of motors that distribute the contents of cells, the proteins of membranes that allow a cell to sense its environment.

It is a truism that the events of biology play out in time and time is the dimension hardest to resolve at the molecular and cellular level. It has felt like Heisenberg's uncertainty principle was at work: the better the spatial resolution, the less we know about the dynamics. In this arena, we are seeing the beginnings of a true revolution, with tools emerging that can give us spatial resolution measured even in nanmometers and still allow us to watch molecular events unfold. We can hope that the detailed events of biological catalysis will grace our textbooks in coming editions. Other techniques allow even single protein molecules or aggregates to be followed as they perambulate inside cells.

We must acknowledge that it is not only proteins that make up cellular life: lipids and carbohydrates — remembered from our textbooks but put aside by most researchers because they did not fit into the central dogma (even as modified by reverse transcription) — have stories to tell of specificity, control and interactions and these are just beginning to emerge.

They will surprise, with that surprise coming to us soon because they are drawing 21st-century attention.

What moves through the world is not proteins or carbohydrates or lipids: it is an organism, an integration of the molecules of life. In this early decade of the 21st century, there are stirrings of interest in integrative biology, usually under the rubric of systems biology and for metabolic questions, this, too, will yield over time to our computers, models and data collection. It will leave us with the toughest problem of biology facing us as the last frontier: the nervous system. Here we face integration on its largest scale, but an integration that can produce the ultimate emergent property: consciousness.

While cellular life outside of the nervous system will be unraveled and known in exquisite detail in the next few decades, the understanding of brains seems likely to tantalize biologists well into the century before a deep understanding emerges. But emerge it will and there will one day be sense to the obvious but presently incomprehensible statement that consciousness is the product of cellular behaviors.

And what will that leave us? What will be our concerns as the century ripens into older age? There will be concerns about synthesis. Just like chemists, we will prove that we understand biology by synthesizing it. And we will not just make over what evolution has produced through its slow, cumbersome four billion-year-long channeled meanderings, we will do it better. We will make things that evolution never got to. The power of gene manipulation will be not only to correct defects but also to create new capabilities. We will not be constrained by the usual constituents of cells. Carbon nanotubes and other nanostructures will interface with proteins providing new levels of control and readout. Medicine will not just be about curing but about replacing and ultimately enhancing.

As we come to the latter decades of our new century, our world will be very different. We will have harnessed the sun's energy directly, replacing the burning of stored chemical energy and watching a hot earth begin to cool back down: letting the universe slowly scour us of our poisonous gases.

Biology will have played a role there, somehow being the model for the capture and storage of the sun's energy. That will have been the conquest of the century. We will live long lives, battling disease to a standstill when it appears. We will lead enhanced lives, tightly linked to computers that will extend our range of activities and do much of the drudgery we now take on ourselves. We will still be humans of the same sort, fighting among ourselves but hopefully not with the guns that will have become so smart that no one can hope to outsmart them.

And we will continue to see vistas. Science will not come to an end because synthetic science knows no limits. There is always another structure to be made, another linkage to be created, another capability to add to our armamentarium. We will still be humans, still restless, still excited by the unrealized opportunity and committed to newness.

At least, that is the world I would hope to leave for those who come after me.