EPILOGUE: Unsolved Mysteries and the T4 Paradigm

BRUCE M. ALBERTS

As a network of many thousands of interacting macromolecules, a living cell is enormously complex and sophisticated. Although the tools that we have for working out the puzzle of the cell have become more powerful each year, they seem completely inadequate for solving the problem. T4 bacteriophage served as a model for one leap in our understanding of living systems forty years ago, when the major tools available were radioisotopes and centrifuges (Hershey and Chase, 1952), and I shall argue that forty years later—in an age of recombinant DNA technologies, massive parallel processors, and whole-genome DNA sequencing—we still need this bacteriophage as a model for understanding the fundamental processes in cells.

When I was a student learning biology thirty-five years ago, we had a view of the cell that was very unlike our view today. We appreciated the powerful catalytic activity of enzymes, but most of us thought of the cell as a test tube crammed full of a random mixture of thousands of these catalysts, between which multitudinous substrates rapidly diffused to be processed into biologically useful molecules. We knew of structural proteins, of course, but we mostly viewed these as components of specialized tissues, such as the actin and myosin of muscle or the keratins of feathers and hair. Today, cells are still complicated, but they make much more sense. Instead of individual proteins floating about, cells are now known to be composed of hundreds of very sophisticated protein machines: complexes of proteins with parts that move in coherent ways, many of them powered by nucleoside triphosphate hydrolyses (Alberts and Miake-Lye, 1992). These machines are, in turn, spatially organized as interacting sets, so that the cell can be thought of as analogous to a very elaborate factory.

In our current view, most individual protein molecules are pieces of large assemblies, and their functions can only be deciphered in the context of the additional proteins with which they normally interact. From this point of view, the cell is like a giant puzzle, and much of it still makes little sense because we are missing so many of the pieces.

Before we can obtain a very deep understanding of biology, we need to be able to describe and analyze whole living systems, rather than just some of their pieces. This is what makes the simplest bacterial cell, like a mycoplasma with its 500 to 1,000 different proteins, so appealing as an object for intense dissection and study. But 1,000 interacting components still represents enormous complexity, and so we must look today for even simpler models to help guide our understanding of the cell. I believe that bacteriophage T4 is just such a model. With its 300 different proteins, and many genes organized on the genome in interacting sets, we have before us an elegant puzzle that has evolved for billions of years. This virus is large enough to carry out many of its central functions, such as DNA replication, genetic recombination, and morphogenesis, with nearly complete independence from host proteins. Thus, in its genome of nearly 169,000 bp, bacteriophage T4 has, in effect, "cloned" out a set of genes that encode the interacting protein machines that deal with selected aspects of metabolism. It is the set of genes that deal with DNA metabolism—DNA replication, recombination, repair, and transcription—that have particularly fascinated me.

As a model of living systems, T4 has many important technical advantages besides its simplicity. It can be very easily grown in the large quantities desired by biochemists, manipulated by classical genetics, and—now that its genome sequence is complete—reengineered at will by reverse genetics. For example, a gene can be cleanly deleted by synthesizing long oligonucleotides designed to contain the gene's two flanking sequences, which can be conveniently delivered into the bacteriophage genome via an insertion-substitution vector system (Selick et al., 1988; Benson and Kreuzer, 1992a; see also chapter 44.6). Moreover, to explore the protein machines of the bacteriophage, one can overproduce nearly any desired protein in an *Escherichia coli* expression vector, attach the purified protein to an insoluble matrix, and use this matrix to fish out the other T4 proteins with which it interacts (Formosa et al., 1991). By using combinations of the above techniques, one should be able to assemble the protein machines that are responsible for a particular bacteriophage function, mutate each of the genes that contribute to this machinery, and decipher the machine's function through combined in vitro and in vivo studies with relative ease.

I have been particularly intrigued with the machine that replicates the T4 chromosome. Although we know its core components well (chapter 5, this volume), it is clear that we still have a great deal to learn about the additional devices that enable this machine to work effectively in the cell. In addition, we are just beginning to explore how it interacts with all of the other protein machinery on the DNA, such as the transcription apparatus (Herendeen, Kassavetis, et al., 1989; Herendeen, Kassavetis, et al., 1992; Liu, B., et al., 1993) and the protein complexes that catalyze genetic recombination (Formosa and Alberts, 1986a; J. Barry and B. M. Alberts, unpublished data). It is already clear that a piece of one protein machine can have functions in a second machine, thereby linking two processes like replication and transcription, which we usually think of as distinct (Herendeen, Williams, et al., 1990a). One would guess that this type of linkage is the rule, rather than the exception. We often tend to forget that each process in a

cell has had to evolve within an elaborate context of other reactions that share the same spaces and the same substrates; this is particularly obvious for the myriad of reactions that occur on DNA.

Why should one study T4 bacteriophage in an age when human biology is directly accessible at the molecular level? My answer is that the shortest distance between two points in biology is often not a straight line. Solving the many mysteries that remain concerning T4 bacteriophage is certain to increase our understanding of all of biology, including that of humans. This has already been demonstrated for the T4 replication apparatus, which we now know to be a simpler form of the apparatus operating in human cells (Tsurimoto and Stillman, 1990). Many of the details concerning the core of the human replication apparatus remain to be deciphered, and it will be many years before we understand its function well. Meanwhile, we should be proceeding to use the T4 replication apparatus to work on the next level of problems, which cannot even be approached today in humans. For example, we can devise more sophisticated in vitro assays to test for the subtle details of replication fork movement, and, once we do this, I suspect that we will begin to find functions for some of the many small proteins now known to be encoded in the T4 genome. Eventually, these same types of proteins will certainly be found in humans, and T4 can provide the guide we need to find them.

For any young investigator about to begin a new laboratory, even T4 must seem terribly complex. Given the wealth of information available from the genome sequence and the power of the tools available to study this bacteriophage, where should one begin? My answer would be to start with one of the deep mysteries that presently confront us in T4 biology. As one example, there is the perennial mystery of what the T4 RII protein does. This famous protein, which was so instrumental in the working out of the genetic code (Crick et al., 1961), is required for a form of lysis inhibition. Like several other mysterious T4 proteins, it is thought to associate with the bacterial plasma membrane following T4 infection. Could its existence be a hint that some electron transport process in this membrane is specially harnessed by the bacteriophage? The fact that we have no clue about how the RII protein might work suggests that we are missing something profound.

In a similar vein, what is the biological reason for the amazing observation that the vast majority of the T4 genes that are transcribed at early times of infection use one strand of the DNA helix as the template for RNA synthesis, whereas most of the late genes use the other strand? There must be something very important that we have been missing here. And there is the perennial question of why the T4 chromosome contains so many genes that have been highly conserved during evolution (as witnessed by their presence in other T-even phages) that can be deleted from the T4 chromosome in laboratory experiments with no effect on phage growth. Many of these "dispensable" genes are clustered on the chromosome, so that large sections of the T4 chromosome can be deleted without affecting phage viability. This clustering might suggest that these genes can move as a "cassette" between one virus and another, and it raises the possibility that they are important for growth of the virus under special conditions that are frequently encountered in nature, e.g., for growth in selected hosts or under anaerobic conditions. It is also possible that only a relatively small growth advantage is imparted by such genes, but that a small effect is sufficient to explain their evolutionary conservation. This is a problem of great general interest, inasmuch as a major proportion of the genes in all organisms seem to be similarly nonessential under standard laboratory conditions.

In summary, the publication of this book on bacteriophage T4 and the simultaneous completion of the entire T4 genome sequence should catalyze a new round of attention to this organism, which has yielded so many important discoveries in the past. For young scientists, all that is required is a careful focus on an important biological problem that can be illuminated by making use of this wonderful biological resource. In these times when almost any experiment is possible, the first major problem is to decide exactly which nugget to mine; a second, equally difficult problem is to develop the judgment required to avoid the 99.9% of experiments that represent distractions from this focused goal. Good luck!