Introduction

THIS BOOK WAS WRITTEN FOR SCIENTISTS of all kinds. It is unapologetically historical. But, you say, the world of biology is rocketing ahead at a pace undreamed of even a decade ago. The advancing technological age in biology that began roughly 35 years ago with the "recombinant DNA revolution" now presents a daily mountain of new information. So why be so misguided in the midst of this whirlwind of the new as to turn out a history? And why a history of RNA?

Consider this: How in the next couple of decades are newcomers to biology going to learn, and how and what are established scientists going to teach them? Already, virtually all college-age students have had exposure, often since grade school, to the mantra "DNA makes RNA makes protein." In this computer age, the notion that biology is an information science and that DNA is the library seems a congenial concept to most who are inclined toward an analytical/scientific career. Perhaps the sensible and necessary course to properly prepare declared biology students and analytically trained "transfers" (mathematicians, physicists, engineers) is first to serve up a predigested catechism of settled conclusions achieved in the 20th century by "wet" laboratory experiments. With this concise biological "periodic table" under command, the newcomer then can be efficiently prepared to deal with the rapidly advancing technology both for doing experiments and for collecting and analyzing to a useful purpose the enormous quantity of data that emerges from today's genomic, proteomic, and computationally enhanced microscopic investigations.

It is by no means my intent to deflect teachers/scientists (mostly young, under 40 years of age) who must carry out the indispensable task of getting students ready to enter today's biology world. Rather, my aim in writing this

book is to provide a supplement in historical form—both to the younger generation of scientists and teachers and through them to incoming students—that describes how we first learned some of the molecular fundamentals of biology in the days of the "hands-on wet laboratory."

One can legitimately argue whether a 2011 biology student "needs" to know any pre-1990 history. I am not prepared to defend vigorously the affirmative in this debate. But I will argue that many *may choose to know* how we came to know all that we did in the era before commercial kits and genomic sequencing took hold. Many of today's major questions (e.g., about how messenger RNAs [mRNAs] are formed and about how gene control is exercised in eukaryotes) are the same questions that were pursued in 1962–1980. More detailed answers to these questions are arriving today at breathtaking speed, but fundamentally informative and important answers came between 1962 and the early 1980s. How these still-central questions first arose, and how early experiments were structured and answers obtained, it seems to me, ought to be at least available in usable form for teachers and, most of all, the curious students of today.

I have been privileged to listen in on a number of "after hours" (read "faculty club cocktail hour") discussions among physicists. Both elders and youngsters in that community seem able to discuss where ideas, questions, and answers came from, easily back to Maxwell and his equations. Biological science, specifically the role of RNA in current and past life on this planet, also has a history worth knowing, I believe.

This history begins with the following questions: How did macromolecules finally become recognized as the necessary starting place for first learning about biology and now teaching biology? And why was RNA the latecomer in this overall picture?

The bold discovery by James Watson and Francis Crick of the structure of DNA, often told, and well told, by the protagonists themselves, is frequently recited as the "start" of molecular biology. And if one watershed discovery is to be chosen as the "beginning," that discovery is it. But there was a preceding half-century struggle of genetics and physical biochemistry that prepared at first a small group of scientists to grasp what the Watson-Crick structure both predicted and demanded but did not answer. For the discovery of the DNA structure and following discoveries to occur, biology/biochemistry had to take on *macromolecules*. The reign of organic chemistry (and recalcitrant organic chemists) as the main route to understanding life had to be at least momentarily sidestepped so that large molecules, poorly understood and comparatively difficult to study, could become the major research focus. How molecular biology involving macromolecules emerged from these early 20th-century battles is fascinating history.

The double helix discovery instantly revealed, through what Crick in his book *What Mad Pursuit* called "such a beautiful structure" (p. 60), how the molecule worked in inheritance. But the Watson-Crick revelation also lit the fuse that led to uncovering the centrality of RNA to life. The miracle years of 1955–1961—just 50 years ago—finally saw RNA recognized to be not monolithic but a collection of different types of molecules with specific functions. Courtesy of the insight of François Jacob and Jacques Monod, biological specificity among different cells, formally a completely opaque problem, could now at least provisionally be explained by controlling the synthesis of specific mRNAs.

The establishment of RNA function—first by discovering how genetic information is transferred into a readable form and then by proving the intimate roles of RNA in translation—led shortly to deciphering the universal genetic code, the first breakthrough toward which Marshall Nirenberg carried out in 1961. But all of these heady achievements were accomplished (largely) with bacteria and their viruses.

As biologists took these ideas to eukaryotic cells, first with cultured human cells, RNA remained the major focus. Throughout the 1960s and 1970s, a new world of macromolecular genetics was unearthed through studies of eukaryotic RNA. Unknown at the time, storage of information in the DNA of eukaryotes was very different from that of bacteria. Simply copying DNA into RNA did not suffice for genetic function. Primary RNA transcripts required molecular carpentry of various kinds—generically termed *RNA processing*—to produce functional RNAs. This era culminated in 1977 with the discovery of pre-mRNA splicing to produce functional mRNA. Both the complicated machinery for digging out the primary transcript as well as the processing to make a specific mRNA opened our eyes to additional points at which regulation of mRNA might occur. All of this was well established *before* facile genomic sequencing confirmed these conclusions.

Soon thereafter (1979–1981), a second bombshell burst. Chemical catalysis can be performed by pure RNAs, most often held in the proper tertiary structure inside cells by protein scaffolds. These major new concepts largely dealing with making and controlling functional RNAs also preceded the era of rapid DNA sequencing.

The young student of today or their youngish mentors can hardly be blamed for knowing very few details of this era, which ended before they were born (in case of students) or before they had finished their first decade or had begun their college years (in the case of young professors). The intellectual sweep of these many achievements before the early 1980s would, of course, be available by reading a selected sample of the hundreds of original papers from 1960 to 1980. But a relatively abbreviated historical discussion

told from the point of view of a long-interested RNA biochemist has been unavailable. This is what compelled me to assemble the material in this book. I note here that I began working with the RNA of poliovirus in the late 1950s in the laboratory of Harry Eagle at the National Institutes of Health. This was followed by a year (1960–1961) with François Jacob at L'Institut Pasteur at the time when the concept and proof of mRNA were just being described (although I had absolutely nothing to do with these landmark experiments). However, my own laboratory work was, and still is, directed by these early very fortuitous training experiences and perhaps will help the reader to forgive the personal voice that appears in various spots throughout the book.

Chapter 1 presents early discoveries that were not fitted into an understandable fabric of cell function for decades. For example, more than 100 years ago, organic chemists were able to identify all of the nucleobases, even placing uracil only in "yeast" nucleic acid (aka RNA) and thymine only in "nuclein," later "thymus" nucleic acid (aka DNA). However, only in 1920 was deoxyribose finally identified as the sugar in DNA.

The peptide bond was described and accepted as the most probable link among amino acids by 1902. But a long disputatious history of the molecular nature of proteins followed, literally until after World War II. How could scientists of 1950 begin to think of cells making proteins by uniting amino acids in the correct order (step by step and therefore uncovering RNA functions) until after 1951, which brought Linus Pauling's models of the α-helix and Fred Sanger's sequencing of the first chain of insulin?

The monumental accomplishments of George Beadle and Edward Tatum in showing that genes were responsible for the function of individual proteins (enzymes) and the discovery of DNA as the genetic material by Oswald Avery, Colin MacLeod, and Maclyn McCarty are stories that preceded Watson and Crick and are known by many at least in outline. But a recitation of exactly what experiments these heroes performed does not trip lightly off the tongue of the majority of today's biologists, young or old.

Therefore, in diplomatic language, after "frank discussions" with my editors, and with the support of many colleagues who have read early versions of the book, Chapter 1 presents some of this history of the centrality of macromolecules, with my hope that, at the very least, it will be entertaining.

Chapter 2 needs no such defense. If we were going to have a history of RNA, it was obligatory to recount the signal achievements of the 1950s–1960s that finally brought RNA out of the shadows. On reflection, viruses with only RNA as a genome were obvious candidates to first catch attention for the genetic/biochemical importance of RNA. This proved to be the case,

with TMV (tobacco mosaic virus) and the RNA formed after T-even (T2, T4) bacteriophage infection of *Escherichia coli* leading the way.

Although the gene/protein connection was made in the 1940s, it took in vitro protein synthesis by rat liver extracts, largely carried out by Paul Zamecnik and colleagues, to begin to truly connect proteins and RNA in 1953–1958. These years saw the discovery of transfer RNA (tRNA) and established a role for ribosomes (and presumably their RNA) in making proteins. Given these advances, it remains something of a puzzle as to why it took so many remarkably gifted scientists so long after the Watson-Crick structure (~7 years) to hit intellectual and molecular pay dirt with the idea of and discovery of mRNA. This is one of the most intriguing stories in the history of molecular biology. The secret lay in closer attention to the genetics of gene regulation that explained switches in the proteins that the cell made. Making mRNA in a controlled fashion in the test tube was then accomplished with bacterial systems in the late 1960s and early 1970s. Discussion of all these accomplishments constitutes Chapter 2.

Chapter 3 guides the reader stepwise through achievements that unlocked the universal genetic code. Virtually all biology students and many from other disciplines will know the conclusions of this era. The aim, however, is to put some experimental meat on the bones of the catechism symbolized by "DNA makes RNA makes protein." The clarity of these conclusions led Gunther Stent, a physicist turned biologist, to title a 1968 paper in *Science* "That Was the Molecular Biology That Was," in which he seemed to argue it was all over but the shouting.

The stunning achievements on bacterial gene functions *were* the logical takeoff point for beginning to work toward understanding how eukaryotic cells controlled their genes and performed such tricks as differentiation. Chapter 4 picks up this problem that began, however, with several years of inability to define mRNA in eukaryotic cells. Kinetic studies following incorporation of labeled RNA precursors into properly separated classes of RNA in cultured animal cells and in animal cells infected with DNA viruses finally wrestled this problem to the ground. Processing of preribosomal RNA (pre-rRNA) was uncovered by 1962–1963 and processing of pre-tRNA by 1968, but RNA processing to make mRNA was not completely understood until the final details of splicing of adenovirus pre-mRNA into mRNA were discovered in 1977. This ~15-year period (1962–1977) is recounted in detail in Chapter 4. Also described in this chapter are the initial discoveries of the biochemistry of the three eukaryotic RNA polymerases and the original illustration of the complexity of nuclear factors required to initiate RNA synthesis correctly in cultured human cells and cell-free systems.

Between the discovery of RNA splicing in 1977 and the early 1980s, additional astonishing discoveries occurred. The capacity of RNA to perform enzymatic functions was recognized. Also, the involvement of previously unknown small ribonucleoprotein particles (in particular, their RNA) in carrying out splicing to make mRNA was discovered. These topics are also introduced in their appropriate historical frame in Chapter 4.

Finally, complementary DNA (cDNA) cloning and pulse-labeled nuclear RNA allowed the measurement of the rate of synthesis of individual genes, which clinched the previously widely assumed, but not yet proven, primary control of gene expression at the level of transcription.

The end of Chapter 4 marks a dividing point in the book. A reasonably comprehensive historical accounting of important events in which RNA is the chief actor ends.

Chapter 5 is an attempt to provide a useful summary of important events in work on RNA after the early 1980s. Because the regulation of mRNA is the central event in all biological specificity, a discussion of the arcane array of proteins that control regulated transcription of chromatin is the first order of business. The positively required factors for the initiation and manufacture of an mRNA from pre-mRNA had first to be understood. This allowed more recent proofs of the wide variety of negative-acting proteins and protein complexes in preventing initiation and of the details of differential pre-mRNA processing, also a regulated process.

The most recent stunning advances in regulation of mRNA translation efficiency and lifetime have come from discoveries of yet more and different RNA molecules, both short and long *noncoding* RNAs. How these were uncovered and initial insight into how they function are products of research in the last ~15 years, and new discoveries and insights into noncoding RNAs continue with each new journal issue. A running summary, necessarily incomplete, of all this experimental activity brings up the rear of Chapter 5. No attempt is made (even if it could have been) to be comprehensive in the material of Chapter 5. Rather, important areas are included with discussion of some key discoveries, and up-to-date references are provided. The shelf life of all of these new findings makes discussions about them admittedly problematic. But no attempt to describe a history of RNA could fail to include a digest of this recent material.

Chapter 6 is brief and highlights, first, a research area that looks back on 3 billion plus years to how RNA likely had an indispensable role in initiating life on the planet. Organic chemical and geochemical advances are being made that may enlighten us about events in the Archaean era. This is possibly the most difficult of all areas in biology, but it no longer seems the impossible field that it did a couple of decades ago.

Perhaps of equal difficulty is the tangled problem of the origin of cells. Challenges to the conventional wisdom of prokaryote → eukaryote evolution began with Carl Woese's discovery of archaea more than 30 years ago and remains a fascinating, unsettled area today despite hundreds of genomic sequences of microorganisms. One of the most challenging unsolved problems in this area also centers on RNA. How and when did splicing of RNA arise—before or after cells arose? If we had an unambiguous answer to this question, might we not also better trace how the three cellular kingdoms arose and persisted? The evolution of cells is intimately tied to the idea of an initial *RNA world*, a hypothetical but increasingly probable time in the evolution of life. Given this likely history and considering the many functions of RNA in the cells of today, shouldn't RNA share a wedge of the spotlight with DNA?