3 Halcyon Days

THE BIG QUESTION

Following the double helix, the overarching question was how the information in DNA is transduced into proteins. It was by then established that, in eukaryotic cells, DNA resided in the chromosomes and doubled at cell division. These observations were consistent with its role as the carrier of genetic information, but not with an involvement in protein synthesis.

In the early 1950s, Jean Brachet showed that enucleated cells could temporarily maintain protein synthesis,[1] and based on a series of grafting experiments with the unicellular green algae *Acetabularia*, Joachim Hammerling showed the existence of morphogenetic substances produced "under the influence" of the nucleus and transported to the cytoplasm that are "products of gene action, which stand between gene and character".[2]

RNA gradually emerged as the intermediate. It was found to be present in high levels in the cytoplasm, particularly in association with the "ergastoplasm" (endoplasmic reticulum) structure[3–5] (Chapter 4), and that its levels vary in different tissues and metabolic states. The correlation between the amount of RNA and the rate of protein synthesis, independently observed in the early 1940s by Brachet and Caspersson, led them to propose that RNA was involved in protein synthesis.[5,6]

DISCOVERY OF THE RIBOSOME

Around the same time, using ultracentrifugation to fractionate mammalian liver cells infected with the cancer-causing Rous sarcoma virus, Albert Claude,[a] who was the first to isolate the mitochondrion, the chloroplast, the Golgi apparatus and the lysosome, observed cytoplasmic granules associated with membranes, initially called microsomes. He found that the granules contained large quantities of nucleic acids of the "ribose type",[8,9] which Brachet postulated to be the sites of protein synthesis.[10] Later, microsomes were found to correspond to microvesicles, arising from fragments of membranes from the endoplasmic reticulum, with which ribosomes are commonly associated. The granules were visualized in 1955 using electron microscopy by George Palade and Philip Siekevitz,[11] and re-named 'ribosomes' in 1958 by Richard Roberts,[12] in view of their abundant ribonucleic acid component (Figure 3.1).

Ribosomes were initially characterized by their physical sedimentation rates, with bacterial ribosomes designated as '70S' and eukaryotic ribosomes as '80S', terminology that is still in use today. Subsequently it was found that ribosomes could be separated into two main components, a large '50S' subunit and a small '30S' subunit in bacteria, and equivalently a large '60S' subunit and a small '40S' subunit in eukaryotes.[13] These studies were aided by the use of detergent solubilization, chaotropic agents and phenol extraction techniques to isolate intact RNAs,[14–16] prior to which most preparations contained mainly degradation products due to the ubiquity of RNases released in lysed cells and present on skin.[17]

Small ribosomal subunits were found to be composed of a number of proteins complexed with an RNA termed 16S and 18S rRNA in bacteria and eukaryotes, respectively. The large subunit in bacteria contains two RNAs (28S and 5S) whereas the large subunit in eukaryotes contains three RNAs (30S, 5.8S and 5S), all complexed with proteins. The ribosomal RNAs are transcribed as a single large precursor, which is processed to form the individual rRNAs, from one operon in bacteria and many in eukaryotes, the latter shown later to be tissue- and developmental stage-specific,[18] as are many ribosomal proteins.[19,20]

---

[a] Claude was also the first to show, controversially, that the active agent in the cancer causing Rous sarcoma virus was not "thymonucleic acid" but "strongly positive … for pentoses" (i.e., RNA), 6 years before Avery.[7] Rous sarcoma virus became famous later for its role in the discovery of the first 'oncogene' (Chapter 6).
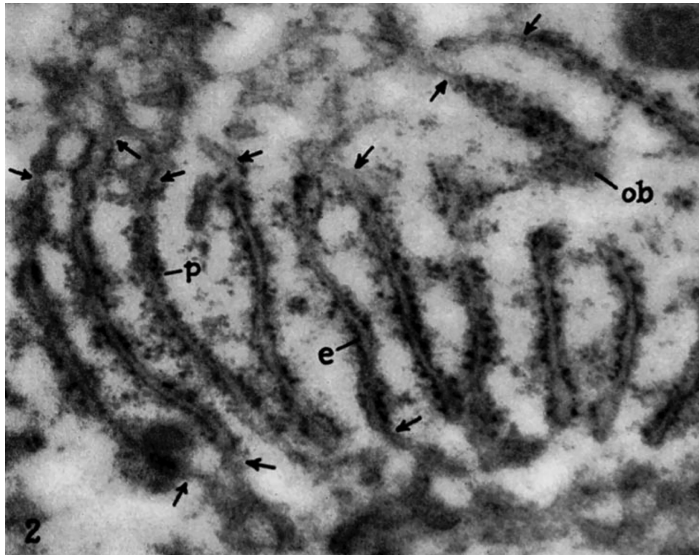
FIGURE 3.1   Electron micrograph of 'ribosomes' associated with the endoplasmic reticulum and free in the cytoplasm of liver cells. (Reproduced from Palade and Siekevitz[11] with permission from Rockefeller Institute Press.)

It was also shown by radioactive tracing that ribosomes are the sites – the cellular factories[b] – where amino acids are assembled into proteins,[21] although the mechanism was yet to be defined.

Thus, the original association of RNA with protein synthesis was a result of the detection of the most abundant RNAs (i.e., rRNAs) allowed by the techniques of the time. This abundance obscured the far more complex population of other RNAs that are expressed from the genome, as later the relatively high abundance of messenger RNAs also obscured the presence of equally if not more complex populations of cell-specific regulatory RNAs in plants and animals (Chapters 12 and 13).

THE MESSENGER AND THE ADAPTOR

The discovery of messenger RNA (mRNA), and its templating of protein synthesis in ribosomes by interaction with adaptor molecules, was science at its best, involving the interplay of observation, logical deductions, discussions and ingenious experiments by many individuals, notably François Jacob,

Jacques Monod,[c] Crick, Watson, Sydney Brenner, Marshall Nirenberg and their collaborators.[24]

The dominant hypothesis that emanated from the 1940s to explain protein biosynthesis was summarized in 1950 by Peter Caldwell and Cyril Hinshelwood: "In the synthesis of protein, the nucleic acid, by a process analogous to crystallization, guides the order by which the various amino acids are laid down."[25]

The importance of protein sequence was reinforced by Linus Pauling's team's discovery in 1949 that the electrophoretic mobility, and therefore the amino acid composition, of hemoglobin is altered in sickle cell anemia, for the first time showing the molecular basis of a genetic disease,[26] with the specific amino acid changes in this and other mutant hemoglobins subsequently identified.[27–30]

It was clear by the early 1950s that the nucleus is the source of RNA[10,31] and that RNA could serve as the template for protein assembly, first proposed

---

[b] As both subunits contain many proteins, RNA was simply thought to be the framework for the machine, until it was shown much later that RNA in fact lies at the catalytic heart of peptide bond formation (Chapter 9).

[c] Both Jacob and Monod were eclectic individuals with interesting histories, including participation in the French Resistance in World War II. Between them, in addition to the 1965 Nobel Prize in Physiology or Medicine, awards included France's World War II highest decoration for valour, the Cross of Liberation, as well as the Croix de Guerre, the Légion d'Honneur and the American Bronze Star Medal. Monod also shared a deep postwar friendship with the writer-philosopher Albert Camus.[22,23]
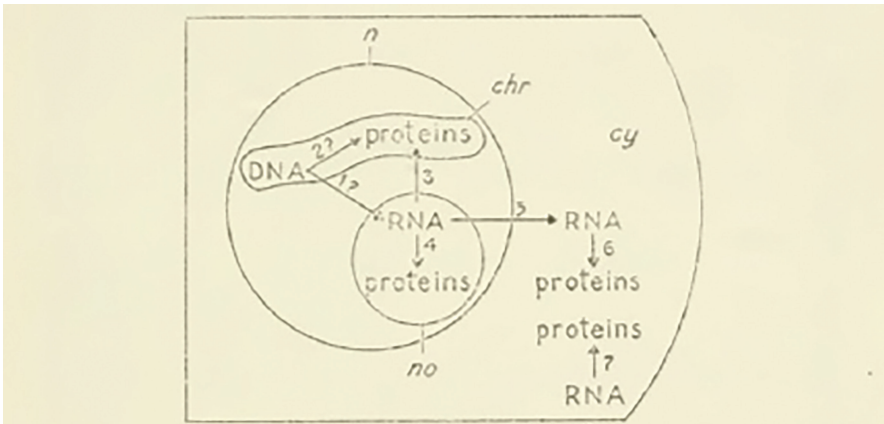
Fig. 41. Scheme proposing relationship between DNA, RNA and proteins in the different parts of the cell. chr: chromosomes; cy: cytoplasm; n: nucleus; no: nucleolus.
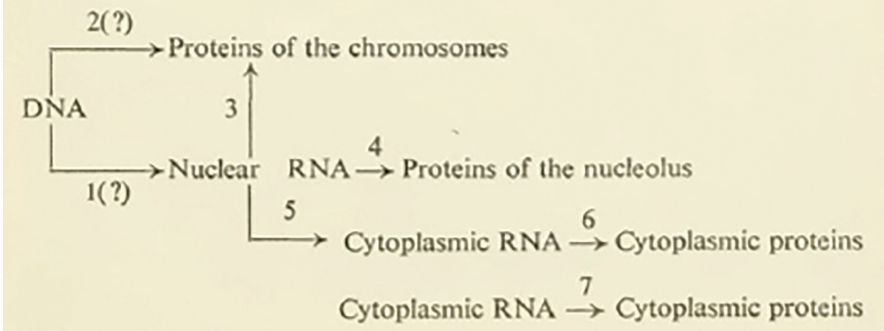
FIGURE 3.2 Jean Brachet's speculations on the flow of genetic information. Figure 41 in his 1959 paper on the biological role of ribonucleic acids. (Reproduced from Brachet,[37] with permission from Elsevier under Creative Commons 4.0 license.)

by André Boivin[d] and Roger Vendrely in 1947[32] and elaborated by Alexander Dounce and Brachet in the early 1950s[33–37] (Figure 3.2). It was also becoming evident that protein synthesis requires "the ordered interaction of three classes of RNA – ribosomal, soluble, and messenger"[38] (see below).

In 1958, as a key part of the theoretical considerations of the process of protein encoding by DNA (the "coding problem"[39]), Crick proposed the 'Adaptor Hypothesis', by which some molecule must serve as the carrier for amino acid incorporation into peptide chains during protein synthesis. Crick postulated that the adaptor was RNA, given that "base

pairing made RNA uniquely suited for a role as a small, specific RNA recognition molecule".[40]

Such RNAs had just been identified by Mahlon Hoagland, Paul Zamecnik and colleagues, who showed that small soluble 'sRNAs' could be conjugated to amino acids (labeled with a radioactive carbon isotope,[14]C) and transfer the labeled amino acids to proteins in microsomal preparations. This reaction required GTP (guanosine triphosphate), later shown to be the energy source for peptide bond formation. From this they concluded that such RNAs, later named transfer RNAs (tRNAs), function as the intermediate carrier of amino acids in protein synthesis.[41]

The factory and the adaptor had been found, but the template and the 'code' remained undefined. With the association of ribosomes with protein synthesis increasingly accepted, one hypothesis was that

---

[d] André Boivin was one of the earliest and most visionary supporters of Avery's claim that DNA was the hereditary material.[24]

distinct ribosomes served as templates for different proteins, leading to the new aphorism "one gene – one ribosome – one enzyme".[40]

However, given the rapid rates of protein synthesis that were observed, for example, after infection of bacteria with bacteriophages (phages[e]) and that the two known RNA species (rRNAs and tRNAs) were essentially homogeneous, stable and similar in different species, this hypothesis seemed implausible: these RNA species did not fulfill the requirements of dynamic templates for protein synthesis.[22,38,45–47]

Early clues for the intermediate candidate had been obtained with the use of nucleotides labeled with a radioactive isotope of phosphorus (^{32}P). In 1953, Hershey and colleagues found that, unlike DNA, a small fraction of RNA is synthesized "extremely rapidly" following T2 phage infection.[48] In 1956, Elliot Volkin and Lazarus Astrachan reported that while T2 phage infection arrested bacterial protein synthesis, it triggered massive phage protein synthesis. They also noticed that while cellular RNA remained essentially unchanged, short-lived RNA with the same base composition as the viral DNA was contemporaneously produced. Volkin and Astrachan called this variant RNA "DNA-like-RNA" and remarked that "such RNA molecules may be an entire new species, possibly related to phage growth".[49] In 1959, Arthur Pardee, Jacob and Monod showed the same rapid inducibility of short-lived RNA from the *lac* operon following lactose exposure[50] (the 'Pajama' experiment[51]).[f]

Others, notably Sol Spiegelman, Benjamin Hall and Masayasu Nomura, confirmed and extended these observations,[52–54] which, although not widely recognized at the time, were crucial for the discovery of the messenger.[22,51] Consequently, Jacob and Monod, in their famous 1961 paper describing the operon model (see below), postulated the existence of an "unstable" RNA that conveyed the genetic information for protein production to the cytosol.

The "candidate" (which they first called "X") was named "messenger RNA" (mRNA).[46]

Brenner, Jacob and Meselson were already working on this hypothesis[22] and in the same year proved the existence of mRNA using the phage system and incorporation of labeled RNA into previously existing ribosomes.[47] At the same time, François Gros, Walter Gilbert and colleagues in Watson's laboratory demonstrated rapid turnover of mRNAs and their "DNA-like" base composition in bacteria,[55] and three groups demonstrated DNA-dependent RNA synthesis in bacteria and in isolated nuclei of mammalian cells.[56–58]

THE 'GENETIC CODE'

In parallel, attention was turning to the nature of the information that instructed the sequence of amino acids in proteins. Theoretical considerations of a 'genetic code' was a major theme in the post-World War II period, especially in the so-called 'RNA Tie Club', which included physicists such as George Gamov and Richard Feynman, as well as Crick, Brenner, Martynas Yčas and others, founded in 1954 to "solve the riddle of the RNA structure and to understand how it built proteins".[59–61]

An intellectual world away from biochemistry but emergent in the background[62] was the revolutionary work in the 1940s on information theory and computational control systems connecting machines with biology by Claude Shannon[g] and Norbert Wiener,[64–66] which ushered in the digital age and, *inter alia*, the concepts of genetic coding and genetic programming.[67]

In 1951, Fred Sanger and colleagues established that proteins are linear polymers of amino acids, and produced the first protein sequence, that of human insulin, by partial hydrolysis of the two peptide chains,[68–71] an approach whose principles he later applied to RNA sequencing (Chapter 6).

Of central importance was Seymour Benzer's use of genetic recombination and temperature-sensitive mutants of bacteriophage T4 at the turn of the decade to map the fine structure of genes.[72–76] Benzer,[h] who later went on to become a pioneer of behavioral genetics,[78] showed that genes are linear but not indivisible, using the resolving power of his system to identify deletions and nucleotide changes, some of

---

[e] The term given to bacterial viruses, from the Greek meaning 'bacteria eater', discovered by Frederick Twort and Félix d'Hérelle in 1915–1917,[42,43] coined by the latter and often shortened to 'phages'. The use of bacteriophages was instrumental in the analysis and elucidation of gene structure, replication and expression, as they comprised an extremely powerful system that could introduce genetic changes and poll millions of genetic events in overnight bacterial culture.[44]

[f] The Pajama experiment also revealed that the induction of beta-galactosidase from the *lac* operon is regulated by a repressor,[50] which ushered in the concept of the regulation of gene expression.[51]

[g] Shannon's PhD thesis was entitled 'An algebra for theoretical genetics'.[63]

[h] Benzer was later described as the researcher who "more than any other single individual, enabled geneticists adapt to the molecular age".[77]

TABLE 3. NUCLEOTIDE SEQUENCES OF RNA CODONS

1st Base	2nd Base				3rd Base
	U	C	A	G	
U	PHE*	SER*	TYR*	CYS*	U
	PHE*	SER*	TYR*	CYS	C
	leu*?	SER	TERM?	cys?	A
	leu*, f-met	SER*	TERM?	TRP*	G
C	leu*	pro*	HIS*	ARG*	U
	leu*	pro*	HIS*	ARG*	C
	leu	PRO*	GLN*	ARG*	A
	LEU	PRO	gln*	arg	G
A	ILE*	THR*	ASN*	SER	U
	ILE*	THR*	ASN*	SER*	C
	ile*	THR*	LYS*	arg*	A
	MET*, F-MET	THR	lys	arg	G
G	VAL*	ALA*	ASP*	GLY*	U
	VAL	ALA*	ASP*	GLY*	C
	VAL*	ALA*	GLU*	GLY*	A
	VAL	ALA	glu	GLY	G

FIGURE 3.3 The genetic code for amino acids presented by Nirenberg et al.[89] (Reproduced with permission of Cold Spring Harbor Laboratory Press.)

which specify a different amino acid and others that corrupt or terminate protein synthesis.[79]

Reasonably, then, genes and proteins were presumed to be co-linear,[80] that is, the order of nucleotides is the same as that of their specified amino acids, but it was unknown whether the code is overlapping or non-overlapping.[i] The former was considered unlikely on logical grounds by Brenner[39] and experimentally by Akira Tsugita and Heinz Fraenkel-Conrat, who showed in 1960 that a point mutation resulted in just one amino acid change.[82]

The matter came to a climax in 1961. Crick and others reasoned that the length of the coding units ('codons') must be at least three to be able to specify all 20 amino acids that standardly occur in proteins, which in turn implied that, if so, there may be more than one ('redundant') codon for each amino acid,[j]

or at least some of them ($4^2 = 16$; $4^3 = 64$), with corresponding 'adaptor RNAs' (tRNAs) linked to cognate amino acids.[40,84]

In 1961, Crick, Brenner and colleagues used Benzer's high-resolution bacteriophage gene system and some of his mutants to show that insertion of one or two nucleotides in the coding sequence resulted in a non-functional protein, because it threw the subsequent codons out of kilter ('frame-shift' mutations), whereas the insertion or deletion of three nucleotides had more subtle effects, thereby demonstrating that the coding unit was indeed a triplet.[81]

In the same year, experiments with RNA homopolymers in cell-free extracts by Marshall Nirenberg and Heinrich Matthaei demonstrated that polyuridine can direct the incorporation of the amino acid phenylalanine into proteins.[85] This not only proved that messenger RNA directs protein synthesis, but also provided the platform for working out the entire triplet-based genetic code by the mid-1960s using combinations of nucleotides in synthetic RNAs[62,86–91](Figure 3.3).

As André Lwoff put it, "the messenger ceased to be an *être de raison* and became a molecule",[92] and the aphorism 'one gene – one enzyme' had found the intermediate.

---

[i] This also led to the common one-dimensional conception of RNAs, which have complex three-dimensional structures, which also transmit information (Chapters 8 and 16). However, only tRNAs were explicitly considered to have a "protein-like structure".[81]

[j] In 1966, following the first determination of the sequence and secondary structure of a tRNA (see below), Crick published 'The Wobble Hypothesis', which provided a structural explanation for the degeneracy of the genetic code.[83]

THE *lac* OPERON AND GENE REGULATION

The year 1961 also saw the publication of Jacob and Monod's classic model of gene regulation in the same paper that proposed the existence of mRNAs,[46] based on studies of lambda phage infection and the genetic dissection of the *lac* 'operon' of *E. coli*. This had a decisive impact on the conceptual framework of the regulation of gene expression and the protein-centric paradigm of genetic information that has dominated molecular biology for most of its history. Every undergraduate student of molecular biology is taught the *lac* operon as the exemplar of gene regulation.

The *lac* operon consists of three 'structural' genes (transcribed as one 'polycistronic' mRNA containing three open reading frames) that specify three proteins involved in the uptake and metabolic utilization of the milk sugar lactose by the bacteria in the gut, including the enzyme beta-galactosidase,[k] together with a nearby 'repressor' gene, whose product keeps the *lac* genes silent until and unless lactose is present – there is no point in producing the enzymes to utilize lactose if none is present.

Jacob and Monod articulated the notion that genomes contained both "structural genes", which encoded enzymes and other proteins, such as hemoglobin and insulin, etc., and "regulator genes",[46] which specified regulatory systems that control the expression of the former.[l] In this model, structural genes obeyed the 'one-gene, one-protein' principle, and regulator genes encoded a trans-acting "repressor" (of unknown composition) that would interact with other DNA sequences ("operators") linked in *cis* (that is, adjacent to the promoter) to block the initiation of transcription, which in turn implied that they would generally lie upstream of the target genes.[m] There was no consideration of the converse, that there may also be activators that operate similarly.

It was discovered later that the product of another gene with a more universal activator function makes it easier for specialized sugar utilization genes to be induced if energy levels are low (Figure 3.4).

In any case, the concept that differential gene activity underlies cell differentiation was obvious and had already been proposed in the 1950s (e.g.,[93]), but whether the regulation of gene expression could be explained simply in terms of the action of regulatory proteins encoded by other genes was uncertain, although beginning to be widely assumed. The cytogeneticist Barbara McClintock intuited that there was a distinction between (protein-coding) "gene elements" and (mobile) "controlling elements", based on genetic studies of transposon mobilization in maize (Chapter 5). She proposed that these mobilized elements, despite not being part of the "gene" nor (likely) enzymes, would act as modifiers, suppressors or inhibitors of gene activity, and predicted their general occurrence in other organisms.[94]

Accordingly, McClintock warned against settling too quickly on a protein-centric definition of genes (and mutations[n]) based on studies in bacteria before the structure of DNA or the nature of the controlling factors she had found – or any 'gene' – were defined. In 1950, she wrote in a letter to a colleague:

> Are we letting a [protein-coding] philosophy of the gene, control [our] reasoning? What then is the philosophy of the gene? Is it a valid philosophy? … When one starts to question the reasoning behind the present notion of the gene (held by most geneticists) the opportunity for questioning its validity becomes apparent.[95]

Moreover, early evidence had indicated that RNAs might have other properties, beyond their roles in protein translation, of potential importance in genetic transactions. Soon after the publication of the double-helical structure of DNA, Alexander Rich (a founding member of the RNA Tie Club) and David Davies showed that RNA molecules could base pair to form double-stranded RNAs (dsRNAs),[96,97] a discovery that was met with some skepticism or disregard,[98]

---

[k] Beta-galactosidase became a favorite target of assays for gene expression by linking its coding sequences to presumed regulatory elements and assaying its activity using an artificial ('chromogenic') substrate that produced a blue color in response to enzyme activity.

[l] And perhaps other regulatory genes, in a chicken-and-egg hierarchy, especially during the complex suites of gene expression during multicellular differentiation and development – see Chapter 15.

[m] What was initially called the 'operator' is now referred to as the gene 'promoter', which encompasses the regulatory sequences upstream of the transcribed regions of genes recognized by regulatory factors and RNA polymerase. Operator may be the better term.

[n] McClintock told Charles Burnham in January 1950: "Even though the details are manifold, obviously, there is a consistency that does not fail. You can see why I have not dared publish an account of this story. There is so much that is completely new and the implications are so suggestive of an altered concept of gene mutation that I have not wanted to make any statements until the evidence was conclusive enough to make me confident of the validity of the concepts."[95]
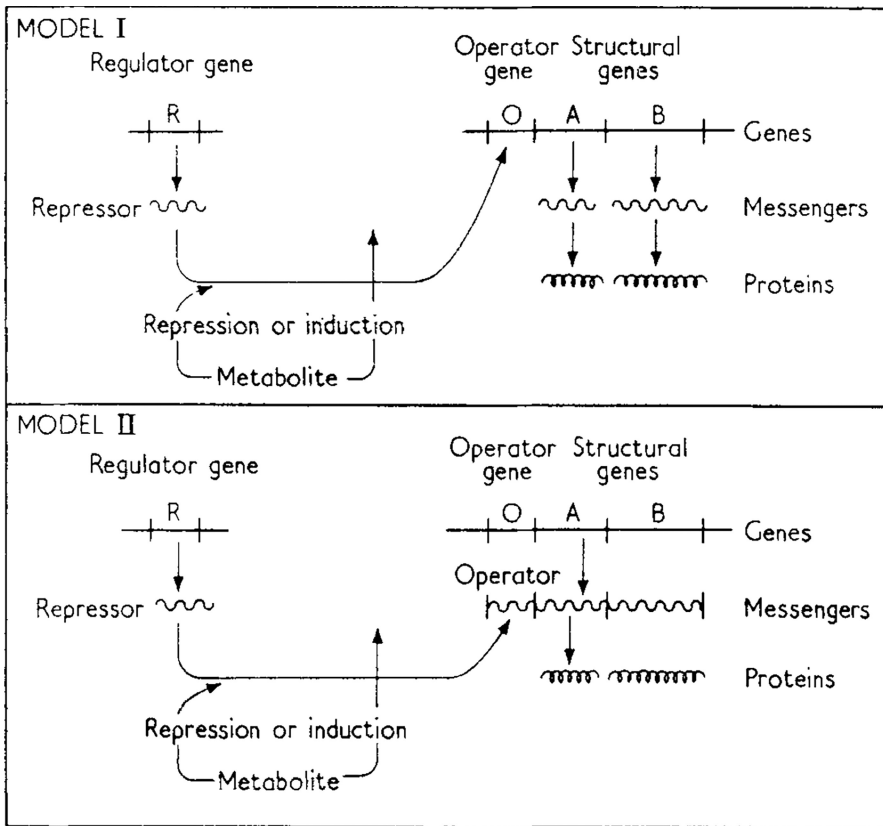
FIGURE 3.4 Jacob and Monod's 1961 models of the *lac* operon and the regulation of protein synthesis.[46] (Reproduced with permission from Elsevier.) Note that in both models the *lac* repressor is drawn as an RNA.

although later shown to be a feature of cellular RNA interactions[99–101] and regulatory systems (Chapters 8 and 12). In 1957, Rich, Davies and Gary Felsenfeld showed that RNA can also interact sequence-specifically with double-stranded RNA or DNA to form three-stranded (triplex) structures,[102] via non-canonical 'Hoogsteen' base pairing[103] in the major groove of the helix,[o] which "may have significance as a prototype for a biologically important three-stranded complex, as, for example, a single ribonucleic acid chain wrapped around a two-stranded DNA".[102] In 1961, Spiegelman and Hall showed that RNA-DNA hybrids exist naturally in cells.[53]

A few years later, Robert Holley, Ada Zamir and colleagues showed by the first sequencing of a tRNA (alanine tRNA, using partial ribonuclease digestion and two-dimensional fractionation[p]) that RNAs form secondary structures via internal base pairing,

forming a 'cloverleaf' structure with double-helical base-paired regions when displayed in two dimensions.[106] These analyses, which took 9 years, also identified ten chemical modifications of its nucleotides[107] (Chapter 17).

The tRNA structure was confirmed and its canonical L-shape revealed almost a decade later by Rich and colleagues using X-ray crystallography, the first determination of the 3D structure of a natural RNA.[108,109] Later studies showed that all tRNAs have four hairpin helices and three variable loop structures inserted between two hairpin structural elements and that the 3′ end of all tRNA molecules contain a conserved CCA sequence, to which the relevant amino acid is attached by specific enzymes (Figure 3.5).[110,111]

These and subsequent structural studies revealed unusual structural motifs, non-canonical base pairing, tertiary interactions, intercalated strands, pseudoknots, coaxial stacking and bound metals, pointing to the structural complexity and versatility of RNA

---

[o] Hoogsteen base pairing occurs in tRNAs.[104]

[p] Fred Sanger and colleagues developed a similar method to sequence RNAs at the same time.[105]

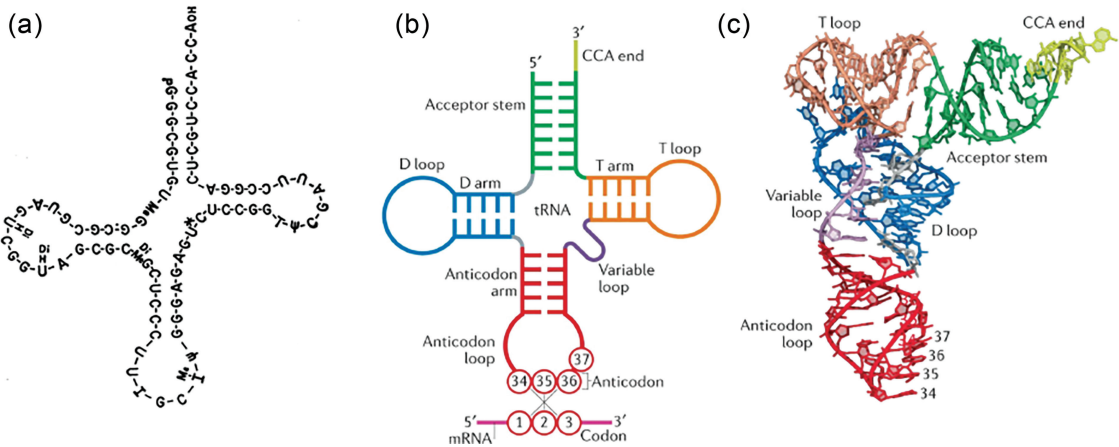**FIGURE 3.5** (a) The original two-dimensional representation of the cloverleaf structure of the 'adaptor' alanine tRNA, drawn by Holley et al.[106] (Reproduced with permission of the American Association for the Advancement of Science.) (b) The general secondary and (c) spatial L-shaped structure (c) of tRNA. (Reproduced from Suzuki[112] with permission from Springer Nature.)

molecules.[108,109,113–116] Although the significance of these properties of RNAs was cryptic,[97] the possible regulatory implications of DNA-RNA and RNA-RNA interactions did not go entirely unnoticed.

In 1959, Arthur Pardee and Louise Prestidge, and a year later Leo Szilard, suggested that RNA would make a good candidate for the *lac* repressor.[117,118] Jacob and Monod subsequently, in their *magnum opus* on the *lac* operon,[46] also proposed that RNA may be the agent produced by the regulatory gene, emphasizing its sequence specificity. In their words: "the operator tends to combine (by virtue of possessing a particular base sequence) specifically and reversibly with a certain (RNA) fraction possessing the proper (complementary) sequence." Because RNA can base pair with DNA and RNA, they proposed two models by which RNA could act as repressors, either at the RNA transcriptional ("genetic operator model") or post-transcriptional levels ("cytoplasmic operator model", where the operator is present in the "polycistronic" transcript). Jacob and Monod favored the former.[46] This was a special moment for RNA in the history of molecular biology, with great conceptual implications, but was short-lived, for reasons explained below.

Rich proposed in 1961 that both strands of DNA in the cells could potentially template complementary ('antisense') copies of RNAs, which was shown much later to be a widespread occurrence, especially in animal and plant cells (Chapter 13). He speculated that "it does not seem likely that both of

these [DNA strands] go on to manufacture a protein molecule" and that there is "an interesting possibility in that this may be part of the control apparatus for turning on or off the synthesis of a given class of proteins", suggesting that this could involve the formation of double-stranded RNA,[119] which also turned out (much later) to be correct (Chapter 12). Again using similar principles, Kenneth Paigen elaborated on the operon model in 1962, and speculated that diffusible (trans-acting) RNAs produced by regulator genes could base pair with the non-template DNA strand of structural genes, reversibly regulating the "release" of messenger RNAs produced from the template DNA strand.[120]

Paul Sypherd and Norman Strauss offered the possibility that the repressor system could involve complexes with both RNA and protein, wherein proteins had specificity for small molecule ligands (in this case, lactose) and the RNA provided specificity for the operator.[121] The latter (RNA guidance of transcription factors and chromatin-modifying proteins to specific genomic locations) was a prescient prediction (Chapter 16). RNA was also later shown to be capable, like proteins, of binding small molecules and responding allosterically ('riboswitches') to regulate gene expression (Chapter 9).[q]

Although it is clear that these models were proposed in the absence of knowledge of many of the

---

[q] Even ribosomal RNA was later shown to regulate the expression of genes that control development.[122]

enzymatic components (such as helicases) and mechanisms involved in DNA replication and transcription, the recurrent theme was that they invoked the "simplicity" and "logic" of RNA regulation via base pairing, which only required an RNA size of 10–12 bases to "provide the necessary specificity",[120] foreshadowing the action of microRNAs and other small RNAs discovered at the turn of the next century (Chapter 12).

Nevertheless, in the years following the *lac* operon, the proposition of regulatory RNAs was disfavored, because the emerging models required that the repressor interact with small molecules (metabolic effectors), for which proteins with three-dimensional structures (such as allosteric enzymes, which alter their shape and activity upon binding small molecules[r]) seemed more suitable, even though these views were "more doctrinal than empirical" and "the proteinaceous nature of the repressor was taken for granted".[126]

This expectation was confirmed and its generality assumed when Walter Gilbert and Benno Muller-Hill found in 1966 that the *lac* repressor is a protein and Mark Ptashne subsequently isolated the bacteriophage lambda repressor protein, both shown to specifically bind to regulatory ('operator') DNA sequences upstream of the protein-coding genes.[127–130] Moreover, in 1965, Ellis Englesberg and collaborators had demonstrated the existence of protein 'activators' in the control of gene expression in bacteria, expanding the dominant 'repressor' model (negative control),[131] although, oddly, Monod was unconvinced.[s] Then, with the discovery of the 'sigma

factors' controlling RNA polymerase transcription initiation,[133,134] the basic mechanism of regulating gene expression in bacteria, and presumptively in higher organisms, seemed generally understood, despite RNA players emerging in the background (Chapter 9).[t]

These findings consolidated the conclusion that proteins comprise not only the 'enzymes' but also the 'regulators' of gene expression, and the suggestions by Jacob and Monod, Rich and others that some genes might specify regulatory RNAs were relegated to history.

PROTEIN STRUCTURE

As aforementioned, an important development during this period was protein sequencing, first achieved in the late 1940s and early 1950s by Fred Sanger using partial acid hydrolysis to sequence insulin,[143] soon supplanted by cyclic cleavage of terminal amino acids called Edman degradation, after its developer, Pehr Edman in 1950,[144] which was later automated.[145]

The determination of the amino acid sequence of proteins (later to be much more efficiently and accurately deduced from gene sequences) was an essential prerequisite to the determination of their three-dimensional structures using X-ray crystallography.[u] It was developed and applied by John Kendrew and Max Perutz and colleagues in 1958 to hemoglobin and muscle myoglobin, which showed

---

[r] The important concept of allostery ('allosteric inhibition') was advanced by Monod and Jacob in 1961 to describe binding of a ligand to one site in a protein causing a structural change that hampers the binding of a second ligand (a DNA sequence or an enzyme substrate) at another site.[123] Monod and colleagues correctly predicted that allostery might be a general form of cellular regulation, and that allosteric sites might be useful drug targets.[123–125] RNAs can also act as allosteric 'riboswitches' that respond to small molecules and other cues, especially in bacteria, not discovered until 2002 (Chapter 9).

[s] During this period, Englesberg gave several seminars at the Pasteur Institute. As told by Jacob: "After each seminar, however, [Englesberg] received a severe lesson in regulatory genetics from Monod, who always insisted on a notion 'that even a schoolboy cannot ignore: negative × negative equals positive!' Englesberg said that 'whenever I spoke with Jacob and Monod, they would say that they were 33.3% convinced, and then 50% convinced, about positive control. When I gave a seminar at the Pasteur … in 1972, they said 'Well, we are 66.6% convinced'."[132]

[t] The identification of the *lac* repressor protein occurred in the same year as the first RNAs not associated with translation were detected in human cells,[135] and just 1 year before the first abundant non-tRNA small RNA was discovered in *E. coli*. The latter was a 6S ubiquitous small (180–200nt) regulatory RNA,[136] whose function as repressor of sigma factor-dependent gene transcription and regulator of RNA polymerase promoter use was not determined until 30 years later.[137] A second small (109nt) RNA found in *E. coli* in 1973[138,139] (a transcript named *Spot 42*, encoded by the *spf* gene) also had an unknown function until 2002: it is also a trans-acting antisense RNA, which represses the galactose operon (and indeed many other operons) at the post-transcriptional level by base pairing with the *galK* mRNA[140–142] (Chapter 9).

[u] The entry of physicists into biology in the 1940s and 1950s revolutionized macromolecular analysis. There is a deeper history, with the development of the principles of X-ray diffraction by crystals and the mathematics involved to probe their atomic structure, dating back to the early 1900s, pioneered by Max von Laue and the father and son team of William Henry and William Lawrence Bragg.[146,147] X-ray fiber diffraction data was, of course, also central to the elucidation of the double-helical structure of DNA.
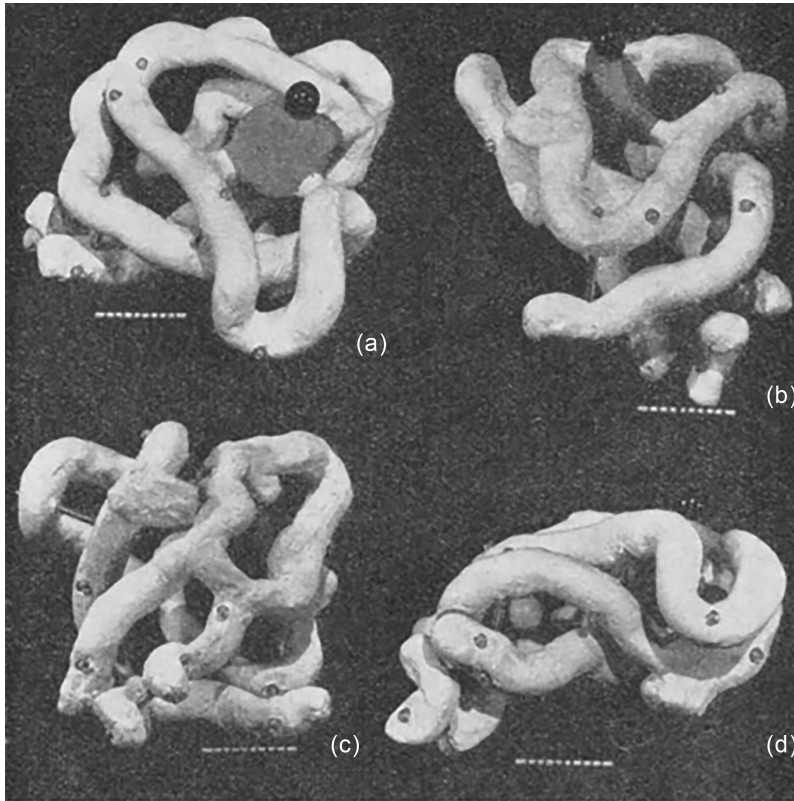
FIGURE 3.6   The first three-dimensional model of myoglobin obtained by X-ray analysis. (Reproduced from Kendrew et al.[148] with permission of Springer Nature.)

that these proteins folded into three-dimensional globular structures[148–152] (Figure 3.6). These studies also revealed the major structural elements of proteins, initially encompassing α-helices, β-sheets, turns and later transmembrane domains and enigmatic 'intrinsically disordered regions' (Chapter 16).

The structural analysis of proteins was initially restricted by their ability to form crystals – creating these was and is an art in itself. Later, nuclear magnetic resonance imaging, first described by Isidor Rabi in 1938 and developed in 1946 by Felix Bloch and Edward Purcell, allowed the determination of relatively small proteins in solution by Kurt Wüthrich, Richard Ernst, Ad Bax, Marius Clore, Angela Gronenborn and Gerhard Wagner, among others, in the 1970s and 1980s,[153,154] and has continued to be refined. More recently, the development of improved methods of cryo-electron microscopy by Jacques Dubochet, Joachim Frank, Richard Henderson[155] and others has allowed structural characterization of much larger proteins and protein complexes.[156–159]

Atomic resolution of protein structure fueled enduring discovery, accelerated by high-throughput methods and many technical innovations that revealed the structure-function relationships, fine chemistry and dynamics in the enzymes, molecular machines and macromolecular components of cells.

THE CENTRAL DOGMA

In the late 1950s, before the identification of mRNA, Crick publicly articulated what he termed the "Central Dogma" of the directional flow of genetic information,[40,160] reflecting earlier considerations by Boivin and Vendrely, Brachet, Watson[v] and Dounce. In this paradigm, proteins were the final destination of the information contained in DNA and conveyed by RNA, as once "information has passed into protein it cannot get out again".[40]

---

[v] Watson sketched the Central Dogma in his lab notebook in 1952.[161,162]
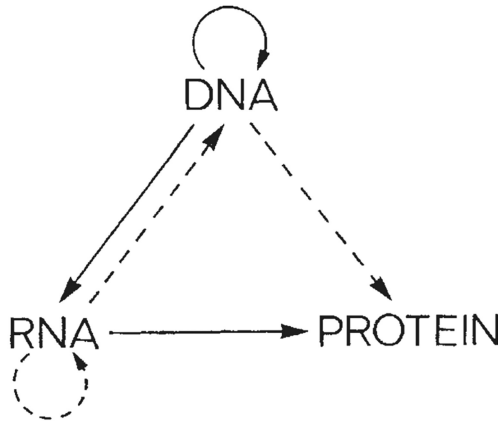
NATURE VOL. 227 AUGUST 8 1970

Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

FIGURE 3.7   Crick's 1970 formulation of the 'Central Dogma'.[163] (Reprinted by permission from Springer Nature.)

In a subsequent formalization in 1970,[163] Crick also included – by way of a dashed line – that RNA may be itself copied, along with a dashed line indicating that information could flow in reverse from RNA to DNA,[w] but not from protein to RNA (Figure 3.7). These modifications were presumably made in the wake of the finding in the 1960s that RNA viruses replicate[164–171] and the discovery of virally encoded reverse transcriptase[x] independently, and with dogged determination in the face of skepticism, by David Baltimore and Howard Temin in 1970.[175–177] Reverse transcriptase was critical to enable the coming gene cloning revolution (Chapter 6) and the understanding of retroviral biology, the full implications of which have yet to be realized (Chapter 10).

The Central Dogma has held true to this day (except for the speculative transfer of information directly from DNA to protein), but became widely interpreted, including by Watson,[162] as 'DNA makes RNA makes proteins', with its implicit

assumption, not necessarily intended by Crick, that RNA functions only as an intermediate.

IT'S ALL OVER NOW

The two decades from 1953 to 1972 were exhilarating and the new crop of molecular biologists were rightly pleased with what had been achieved, but their self-satisfaction and hubris were palpable. The *lac* operon and the Central Dogma consolidated the notion that (with exceptions like the few rRNA and tRNA types) genes are synonymous with proteins, and that all genetic information, including regulatory information, is transacted by proteins, not only in bacteria but also in developmentally complex plants and animals.

Consequently, the hegemony of proteins as both structural and regulatory molecules was established, prematurely, within the first two decades of molecular biology, despite the odd molecular and genetic observations in plants and animals (Chapters 4 and 5) and a looming surprise that should have given pause for thought (Chapter 7), with more to come (Chapters 8–13).

As Crick opined in 1958: "Biologists should not deceive themselves with the thought that some new class of biological molecules, of comparable importance to the proteins, remains to be discovered".[40]

---

[w] Crick's original diagram of the flow of genetic information included a dotted line from RNA back to DNA.[40]

[x] RNA-dependent RNA polymerases are also found in eukaryotic cells.[172] There are a number of 'DNA repair' enzymes with reverse transcriptase activity in the brain, with obvious implications.[173] Information also moves laterally from DNA via RNA to DNA by retrotransposition,[174] the outcomes of which dominate the genome and genetics of complex organisms (Chapters 4, 10 and 16).

And Brenner in 1963:

> It is now widely realized that nearly all the 'classical' problems of molecular biology have either been solved or will be solved in the next decade. … Molecular biology succeeded in its analysis of genetic mechanisms partly because geneticists had generated the idea of one gene-one enzyme. … Molecular biology succeeded also because there were simple model systems such as phages which exhibited all the essential features of higher organisms so far as replication and expression of the genetic material were concerned.[y] … It is probably true to say that no major discovery comparable in importance to that of, say, messenger RNA, now lies ahead in this field.[z]

Gunther Stent proclaimed in his 1968 article entitled 'That Was the Molecular Biology That Was':

> All hope that paradoxes would still turn up in the study of heredity had been abandoned long ago, and what remained now was the need to iron out the details … [and that there remained] … only one major frontier of biological enquiry for which reasonable molecular mechanisms cannot be envisaged: the higher nervous system.[179]

And Brenner added [Stent's point is]

> that once we knew both the structure of DNA and that nucleotide sequences encoded amino acid sequences of proteins, and that once the principle of gene regulation had been found by Jacob and Monod, there was nothing left to do. Thus embryology could be accounted for by simply turning on the right genes in the right place at the right time and that was the solution to the problems of development. Not only did we not have to bother investigating the developmental biology of the millions of different species of animals and plants, but there would be no motivation for scientists to pursue those fields because the mystery had vanished.[180]

The belief that genes are synonymous with proteins reflected the mechanical zeitgeist of the age. Bicycles and cars have parts, and so do organisms – proteins that carry oxygen (hemoglobin), form skin (keratins), signal energy levels (insulin) or control the activity of other genes ('transcription factors'), etc. It was just assumed that these 'conserved' components, whose expression is regulated by *trans*-acting transcription factors acting on malleable adjacent promoter-operator sequences, were enough to explain all of biology.

Little thought was given at the time to the enormous differences between bacteria and developmentally complex organisms. The 'biochemical unity of life' was just taken as given.[181] As Monod said in 1954, in a recapitulation of a 1926 assertion by the microbiologist Albert Jan Kluyver:[aa] "Anything found to be true of *E. coli* must also be true of elephants."[181]

That may be the case, but the logical trap was that the reciprocal might not be. No one knew.

FURTHER READING

Ball P. (2018) Schrödinger's cat among biology's pigeons: 75 years of what is life? *Nature* 560: 548–50.

Cairns J. and Watson J.D. (2007) *Phage and the Origins of Molecular Biology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Carroll S.B. (2013) *Brave Genius: A Scientist, a Philosopher, and Their Daring Adventures from the French Resistance to the Nobel Prize* (Crown Publishers, New York).

Cobb M. (2015) Who discovered messenger RNA? *Current Biology* 25: R526–32.

Cobb M. (2016) *Life's Greatest Secret: The Race to Crack the Genetic Code* (Profile Books Ltd., London).

Darnell J.E. (2011) *RNA: Life's Indispensable Molecule* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Kay L. (2000) *Who Wrote the Book of Life? A History of the Genetic Code* (Stanford University Press, Redwood City, CA).

Morange M. (2020) *The Black Box of Biology: A History of the Molecular Revolution (translated by Matthew Cobb)* (Harvard University Press, Cambridge, MA).

Olby R. (1994) *The Path to the Double Helix: The Discovery of DNA* (Dover Publications, Mineola, NY).

Rich A. (2009) The era of RNA awakening: Structural biology of RNA in the early years. *Quarterly Reviews of Biophysics* 42: 117–37.

Siekevitz P. and Zamecnik P.C. (1981) Ribosomes and protein synthesis. *Journal of Cell Biology* 91: 53s–65s.

Watson J.D., Gann A. and Witkowski J. (2012) *Double Helix, Annotated and Illustrated* (Simon & Schuster, New York).

---

[y] Sydney Brenner, Excerpts from Letter to Max Perutz, June 1963; reproduced in Wood (1988).[178]

[z] Sydney Brenner, excerpts from Proposal to the Medical Research Council, October, 1963; reproduced in Wood (1988).[178]

[aa] "From the elephant to butyric acid bacterium—it is all the same!"[181]