Building a Cell

Jack Bennett

THE CELL IS the basic structural unit in all living things, and cell function is a major unifying concept in biology. Thus, acquiring an understanding of the nature and synthesis of nucleic acids and proteins within the cell is important for science students. To relate many otherwise disparate facts into a coordinated and logical structure, I developed a scheme that illustrates the interdependence of proteins and nucleic acids in hereditary functions. The original scheme, which culminates in one diagram titled the "Hereditary System," has been refined and extended based on the reactions and questions of a succession of students.

The scheme is built around the arguments and assertions on the nature and synthesis of nucleic acids and proteins presented in current texts. My favorite resources are the *Scientific American* articles in a collection by Kennedy (1965) and *The Molecular Biology of the Gene* by Watson (1976). These sources provide suitable background for generating this discussion of the hereditary system. The presentation follows.

DNA Synthesis

First, we review some relevant facts. In figure 1, we see a highly schematic representation of DNA synthesis. For DNA to synthesize, a minimum of four factors must be present; these are:

1. A source of suitably activated deoxyribose nucleotides;
2. A source of energy;
3. A set of enzymes that can connect the nucleotides to make the polynucleotide, hence a DNA polymerase; and
4. A source of template DNA.

As the diagram illustrates, the original DNA molecule, a double helix of two polymers held to one another by hydrogen bonds, separates into two complementary strands. These strands serve as scaffolding for the organization of the deoxyribose nucleotide triphosphates into new complementary sequences. The polymerase can then connect them together, thus completing the process of producing two identical DNA double helices where only one had existed previously.

RNA Synthesis

In figure 2 we see that similar process is involved in RNA synthesis. You will note that the requirements are the same; they are:

1. A source of suitably activated ribose nucleotides;
2. A source of energy;
3. An appropriate RNA polymerase; and,
4. A molecule of DNA to serve as a model.

The DNA double helix separates and serves as a scaffold for the sequential organization of the ribose nucleotide triphosphates, which are then connected into a polyribonucleotide—RNA—by the RNA Polymerase. It does not seem to make any difference which of the several kinds of RNA is produced, the method of production is the same. It is also curious that because of coding signals in the DNA the polymerase is ordinarily only able to assemble the complement to one of the two separated DNA chains. After the polymerization, the RNA separates from the DNA model and the DNA double helix is re-established. Thus, the DNA is returned to its original condition. This process produces Ribosomal RNA (rRNA), Messenger RNA (mRNA), and Transfer RNA (tRNA). The function and activity of each will receive further mention.

In figure 3 we schematize one part of the activity of tRNA. There appears to be a different kind of tRNA for each of the 20 kinds of amino acids incorporated into proteins in ordinary biosynthesis. In fact, there are several specific tRNAs for most amino acids in all organisms studied thus far. The structure of a tRNA is variously represented, but functionally each type of tRNA has several folded regions with an attachment point for the carboxyl of the amino acid and a set of three bases that constitute a code label associated with a particular amino acid. Ordinarily, the amino acids are enzymatically activated by attachment to an ATP or similar carrier molecule and are then transferred to the tRNA by specific "loading enzymes" (called amino-acyl-transferases). The precise and specific attachment of a given amino acid to the appropriate and specific tRNA can only be carried out by the specific loading enzyme for that combination. Thus, we

The author is professor of biology at Northern Illinois University, DeKalb 60115. He received his Ph.D. degree in genetics and zoology from the University of Wisconsin in 1959. Bennett holds memberships in AIBS, AAAS, Phi Sigma, the Society for the Study of Evolution, the Genetics Society of America, and the Association of Midwestern College Biology Teachers. He has published extensively in journals including *ABT, Heredity,* the *Journal of Heredity,* and *Science.* This article is adapted from a paper Bennett presented to the 23rd Annual Meeting of the Association of Midwestern College Biology Teachers held in October 1979.

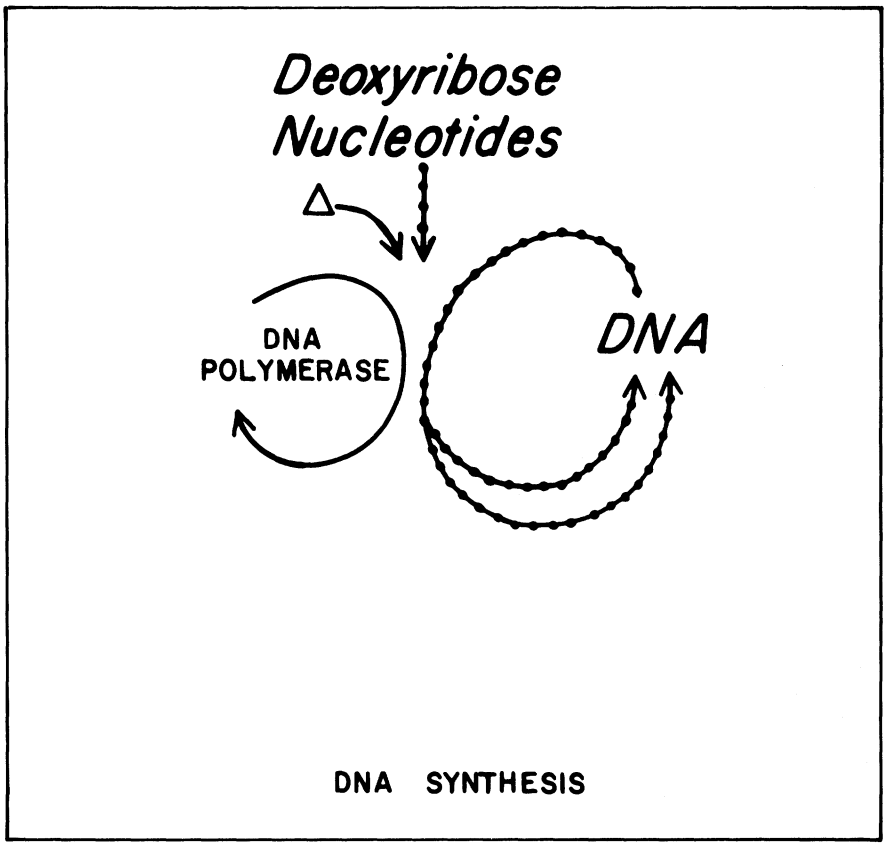

FIGURE 1. The synthesis of DNA requires: The presence of the four deoxyribose nucleotide triphosphates; A source of energy as ATP (adenosine triphosphate); A source of preexisting DNA; and, An enzyme complex, DNA polymerase, to connect together the nucleotides that have been organized on each of the original separated chains.

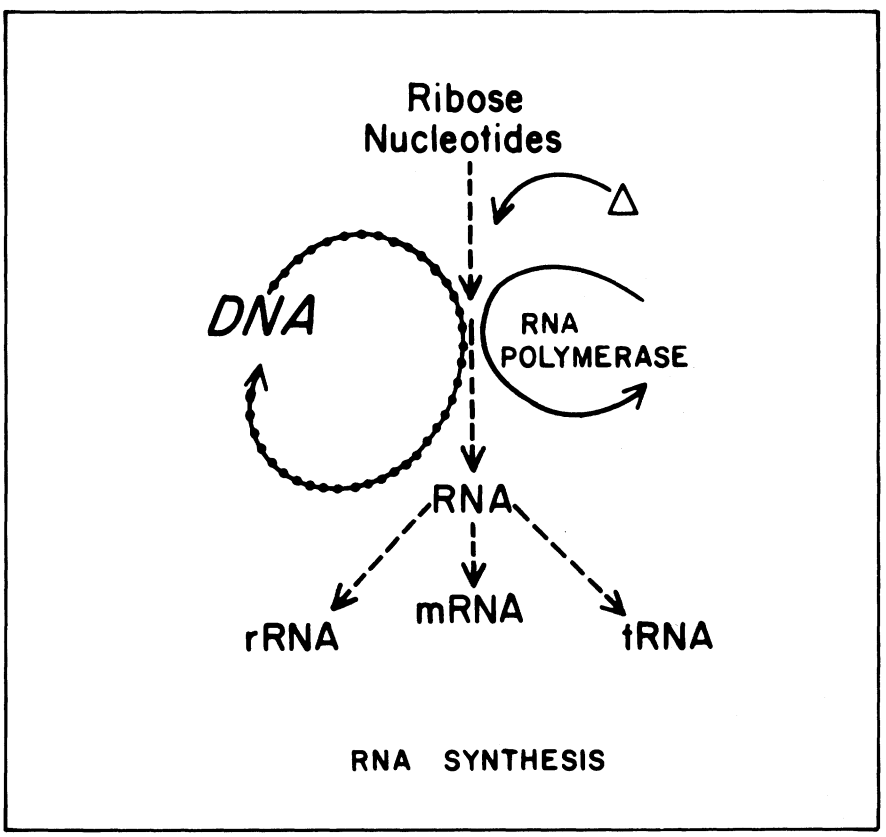

FIGURE 2. RNA synthesis requires: The presence of suitable ribose nucleotide triphosphates; A source of DNA with the double helix separated to allow the ribose nucleotides to be organized along one chain; A source of energy, as ATP; and, an RNA polymerase to connect the nucleotides together to make RNA. The DNA chains recombine afterward into the original double helix and the RNA molecules are released. This process is commonly called transcription.

note that the translation from an RNA "code," or language, to a particular amino acid specification is made at this point by an enzyme the amino-acyl-transferase.

The rRNA, combined with many proteins form a ribosome (fig. 4). This body serves as a meeting place for amino acid carrying tRNAs and the mRNA. The mRNA is effectively a series of code groups—or words in sequences of three nucleotides, along which the tRNAs are ordered on the surface of the ribosome. When the amino acids have been brought into close proximity, and into the appropriate sequence, the protein polymerase disconnects them sequentially from their tRNAs and connects them into a polypeptide chain. These chains fold into unique and characteristic structural forms as they evolve from the surface of the ribosome. They are proteins whose structure is thus another manifestation of the information contained in the sequence of nucleotides of the mRNA. Thus if one knows the linear sequence of amino acids of a protein it would be possible to work backwards and construct a linear sequence of mRNA; and from this information, it is possible to find a complementary sequence in DNA.

Hereditary System

These four representations when combined give us the "Hereditary System" as shown in figure 5. In the combined system, all of the separate processes depend on the others for maintenance. DNA synthesis fails without protein enzyme systems to provide the energy, nucleotides, and the polymerase. RNA synthesis fails without its own similar, but distinct proteins. The supply of amino acids and the loading of the tRNA again requires energy and is dependent on the presence and production of the various RNAs. The ribosome is made of protein and RNA, and its function is clearly dependent on both.

If we were to assemble all of the essential molecules in one place, provide enzymes for making ATP, and enclose them in a membrane to prevent their dispersal we would have an assemblage that bears an amazing resemblance to an old acquaintance: the cell. The cell can take in materials from the environment and synthesize new molecules of its own kind and thus grow. It can use environmental glucose for metabolic energy. When the cell grows sufficiently large it will have multiple samples of each component. Thus it can divide with each product cell having at least one of each essential part. The cell appears to include the minimum collection of parts required to be a growing and self-duplicating system, even though no one of its parts has this property. I repeat: No one molecule or class of molecules can independently duplicate itself. The several classes when organized together in the hereditary system can all be duplicated by their common activities.

This description leads us to consider the complexity involved in the system. Estimating the minimum molecular requirements for each of the reactions in this system is possible. This is feasible for estimates of both the number of different kinds of molecules and of the total number of macromolecules in such a minimum system. If we organize our estimates according to the requirement for each type of molecule, we can build a table of requirements. For DNA synthesis the number of different proteins required (and let us assume that one polypeptide equals one protein in this system), would seem to include the DNA polymerase, at least one enzyme for each of the

TABLE 1. Numbers of Molecules Required for Operation of the Various Portions of the Hereditary System.

DNA Synthesis (Duplication or Replication)	
DNA	?
RNA	-
Protein	14
(4 nucleoside kinases)	
(1 unwinding protein)	
(4 subunits, RNA Polymerase III)	
(1 DNA polymerase/RNA excision)	
(4 subunits DNA polymerase III)	
RNA Synthesis (Transcription)	
DNA	?
RNA	-
Protein	8
(3 ribonucleoside kinases, ATP is already present)	
(5 subunits RNA polymerase holoenzyme)	
tRNA Loading (Charging)	
DNA	-
tRNA	60
Protein	21
(20 amino acyl transferases)	
(1 peptidase (digestive enzyme to provide amino acids from environment)	
Ribosome Function	
DNA	-
RNA	202
(3 rRNA)	
(199 mRNA)	
Protein	62
(54 ribosomal)	
(8 co-factors)	
Energy Production, Etc.	
DNA	-
RNA	-
Protein	84
(1 RNAase)	
(12 Glycolysis enzymes)	
(60 permeases)	
(11 cell division)	
Membrane	
DNA	-
RNA	-
Lipid	4
Protein	10
(4 enzymes for the 4 lipids)	
(6 membrane structural)	

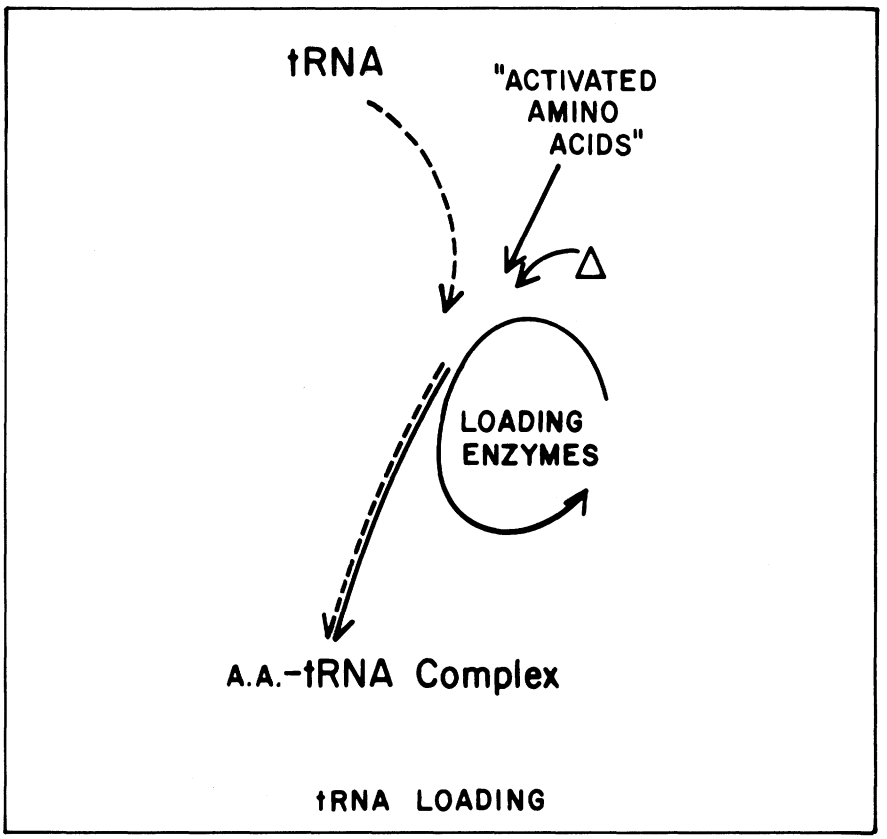

FIGURE 3. A schematic representation of the various transfer RNAs being loaded with the specific amino acids by the specific loading enzymes. These enzymes determine which amino acid goes on to a given kind of tRNA. Thus the enzymes are the actual translators of the RNA "language" into the amino acid "language".

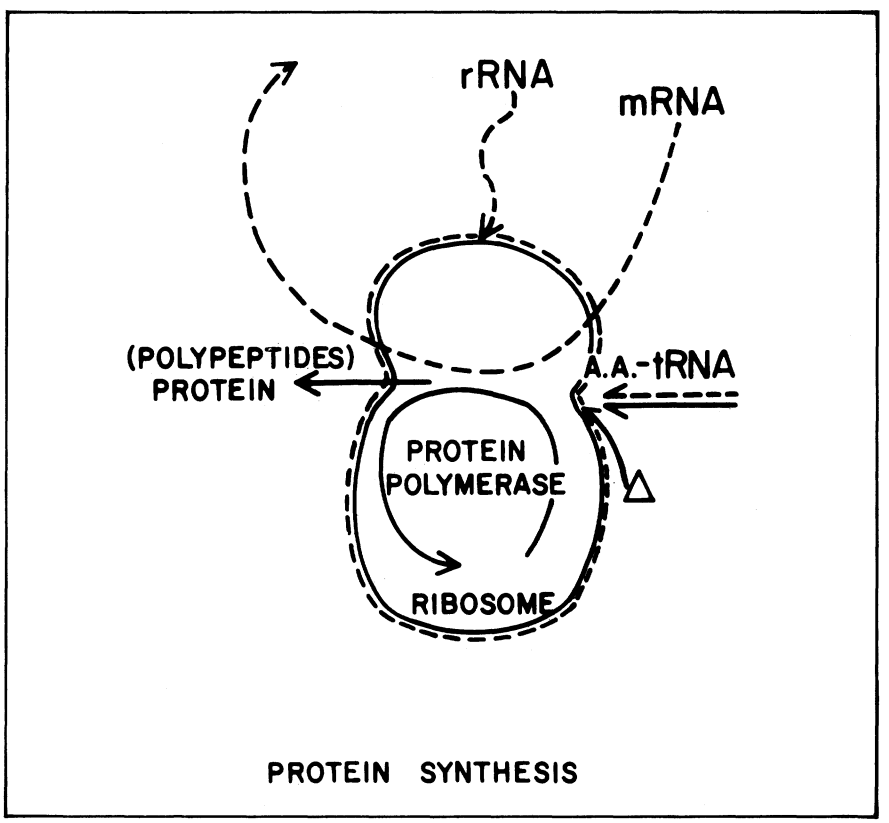

FIGURE 4. A schematic representation of the operation of a ribosome in protein synthesis. On the ribosome the mRNA serves to order the loaded tRNA's. The resulting linear sequences of amino acids are separated from their tRNAs and connected to each other by the protein polymerase enzymes. The "unloaded" tRNAs and the mRNA are released to repeat their activities and the amino acid (polypeptide) chain folds up to become a functioning protein molecule. This process is commonly called Translation.

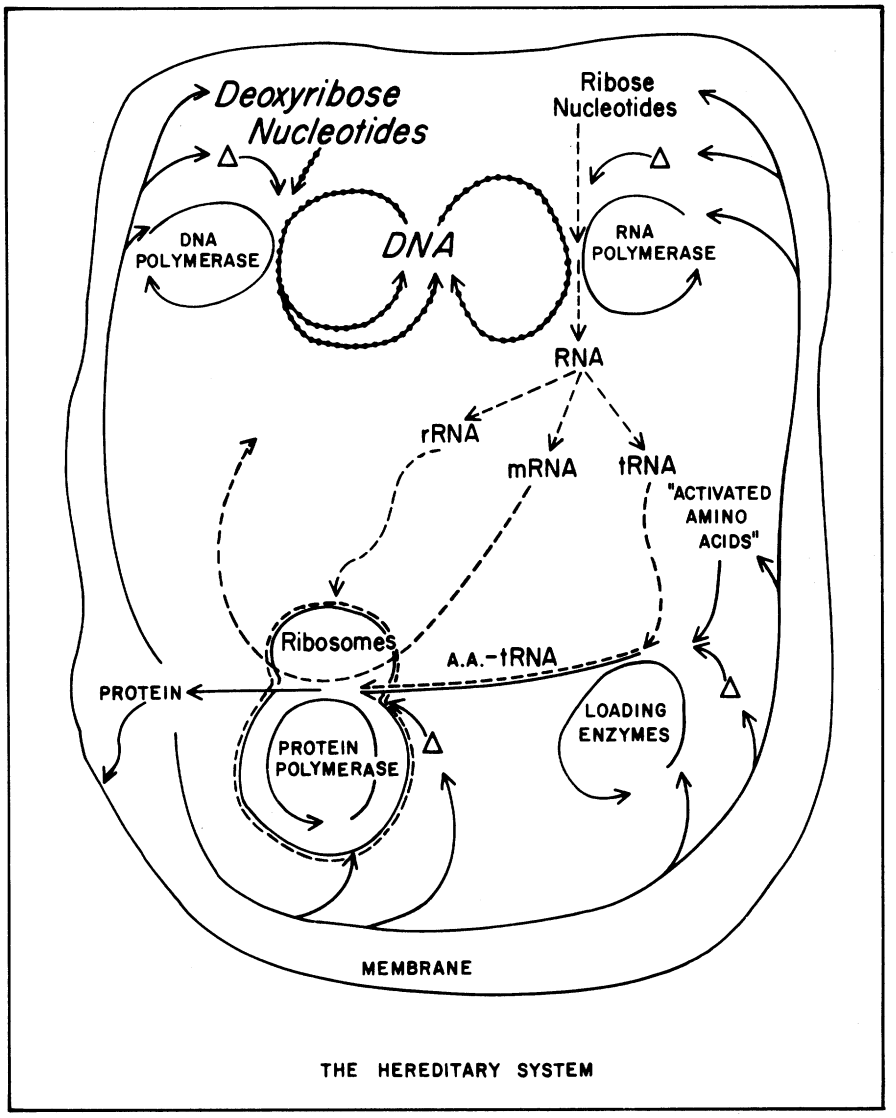

FIGURE 5. "The Hereditary System." A representation of the interdependence of the different processes. DNA synthesis cannot proceed without support of protein enzyme systems and pre-existing DNA. RNA synthesis requires a similar list of proteins and DNA. Protein synthesis itself requires three classes of RNA and the activity of several protein enzyme systems. With the addition of a set of glycolysis enzymes to provide ATP, and a membrane wrapped around it the system bears a startling resemblance to the usual descriptions of a cell. That is, it is an independent unit, it can take in materials from the environment and make more of itself and grow, and thus reproduce.

four deoxyribose nucleotides, and an enzyme system to provide energy. The simplest energy source would probably require twelve enzymes (anaerobic glycolysis, for example). Thus, we can imagine a minimum of fifteen proteins needed for DNA synthesis (table 1). RNA synthesis (and all other energy-requiring processes) will draw on the same energy source as DNA synthesis. A separate RNA polymerase and at least one enzyme for each of the ribose nucleotides seems a reasonable minimum. Thus, we can imagine RNA synthesis proceeding with an addi-

TABLE 2. Summary of Macromolecules for the Hereditary System Model

	DNA Synthesis	*RNA Synthesis*	*tRNA Charging*	*Ribosome Function Etc.*	*Energy*	*Membrane*	
DNA	1 (262)	1 (262)	-	-	-	-	1 (262 regions)
RNA	-	-	60	202	-	-	262
Protein	14	8	21	62	84	10	199
Lipid	-	-	-	-	-	4	4
						Total Number of Different Kinds of Macromolecules	462
Total Number of Macromolecules × Number of Each Kind*							
	141	81	810	849	840	2040	4761
						8000	8000 (membrane lipids)
							12,761

*Assume: 10 ribosomes; 1000 proteins embedded in each membrane layer (plus permeases, etc.); 10 of each enzyme; 10 of each tRNA; 1 of each mRNA; 1 of total DNA.

tional eight different proteins. The loading of tRNA will require a source of amino acids. As a minimum, one might imagine at least one external protease for the purpose. If we assume that a single amino-acyl-transferase can manage to serve all of the tRNAs for each amino acid, then only twenty loading enzymes would be required. Thus, the loaded tRNAs could conceivably be provided by as few as twenty-one different kinds of protein.

Protein synthesis requires the ribosome. A ribosome contains three distinct kinds of rRNA and about 62 different proteins including the protein polymerase.

The cell membrane would require several different kinds of structural molecules in addition to the lipid bilayer—at the least three "inside" proteins and three "outside" proteins. It seems that at least four enzymes are required to produce the lipids. Based on this estimate, the maintenance of the membrane requires at least ten different kinds of proteins.

The requirement for different RNA molecules is closely related to the number of kinds of proteins the cell needs. There appears to be no direct RNA requirement in DNA or RNA synthesis. Sixty different tRNAs are required regardless of the number of different proteins produced. We recall that three different rRNAs are needed. If we add them, as in table 2, we have 199 different proteins and thus would need 199 different mRNAs to code for them. Thus, a total of 262 different RNAs would be needed (table 2).

DNA is required in both DNA synthesis and RNA synthesis. At least 262 different sequences of DNA are needed to produce the 262 RNAs; and, of course, this DNA would also be necessary for DNA duplication.

We have not provided for controls to regulate the activity of the genes relative to each other. In the regulatory systems examined thus far, the number of base pairs of DNA required for operators, promoters, etc., does not approach 10% of the base pairs in the structural genes that are regulated. Therefore, it seems unnecessary to consider these units separately, even though they are formally separate and identifiable sequences.

All of this information is summarized in table 2, which indicates that we need 199 different proteins, 262 different RNAs and 262 different DNA sequences (or regions of one big DNA double helix), or a total of 462 different macromolecules for the system to work.

If we inquire further into the number of each kind of molecule, the result is nearly as satisfying. At least one of the larger viruses has a single layer protein coat made up of about 750 protein molecules. We could imagine our hereditary system being enclosed in a lipid bilayer membrane with a thousand proteins embedded in each surface. Prudence would indicate that the other kinds of protein molecules should be present in multiple samples of each so that loss of a single molecule could not stop the whole process.

The RNAs require a little more thought. Efficient operation seems to require more than one of each kind of tRNA—perhaps 10 would do—thus, we have a total of 600 tRNAs. Perhaps 10 ribosomes would be needed—thus 30 rRNAs. (Note that this requires 620 proteins, also). At least one sample of each mRNA would be needed for a total of 199. We would, therefore, expect 829 RNA molecules. If the system were not in the process of duplication, one of each kind of DNA (one DNA double helix with 262 regions) would be sufficient.

When added, this list indicates that 4,761 macromolecules would be sufficient for a minimum system. An assortment of water molecules, lipids, sugar, ionized salts, and other small molecules would also be present. This would certainly not be the most efficient system possible, but it appears to have all that is necessary to function.

This complexity is clearly within the limits of human comprehension. If we consider the products of human manufacture, we note that many everyday items appear a good deal more complex. A TV set, or an automobile, has more different parts and more total parts by a wide margin than the system we have constructed here. Therefore, we can reasonably assert that the simple form of the hereditary system is more amenable to human understanding than are a large number of human inventions and constructions. We have developed the general specifications for the cell here. The detailed specifications of sequences of amino acids, or the equivalent in nucleotides, are also well within our comprehension. The last volume of the *Atlas of Protein Sequence and Structure* (Dayhoff 1979) contained 366 new protein sequences that had been determined in the preceding 31-month period, more than our model requires.

The theoretical genetic literature of the 1950s reported analyses, based on information theory, that indicated that the minimum number of different kinds of parts for self-duplicating automaton would be of the order of 250 to 400. This total is comparable to the 462 generated here.

I pointed out earlier that the system as developed here is, in fact, a description of a simple cell. If this is so, then we might expect that some organisms might exist with a comparable simple structure. It appears that they do! The *Mycoplasmas* (or PPLOs, for pleuor-pneumonia-like organisms) are plentiful and apparently still successful organisms (Morowitz and Tourtellotte 1962). They seem to be mostly extra-cellular parasites of multicellular forms, though some are free-living. They have only been clearly identified in the past 25 years and their evolutionary history and taxonomic status are still not established. We do not yet have an exact census of the number of different kinds of macromolecules in a given species. However, it appears that some of them go through a portion of their life cycle when they may consist of as few as 1,200 macromolecules as "elementary bodies," which are thought to function as reproductive units. Thus, in their normal environment these creatures seem to be even smaller at times than we imagined they should be!

TABLE 3. DNA Required for the Hereditary System

Proteins: 199 × 500 AA × 3 base pairs/AA		=	298,500 base pairs
RNA: rRNA			
5s	120 bases		
18s	1,818 bases	5,876 bases	= 5,876 b.p.
23s	3,938 bases		
tRNA 60tRNAs x 80 bases		=	4,800 b.p.
Thus, to code all this requires		=	309,176 b.p.

or about 204 million MW DNA, or about 105 microns (= 0.105 mm) DNA double helix.

DNA Content of Certain Organisms Compared

	MW	b.p.	mm
Hemophilus influenzae	800 million	121,200	0.41
Escherichia coli	2,500 million	378,800	1.29
Mycoplasma arthriditis	444 million	67,300	0.228
Hereditary System Model	204 million	30,900	0.105

More recent work on *Mycoplasmas* has provided evidence on the total DNA content of some species. This is available in terms of the molecular weight of DNA per cell. Table 3 presents some DNA molecular weights for several micro-organisms, and for our hereditary system. For easier understanding of the comparative sizes the weights have also been translated in lengths of the DNA double helices.

We have no direct knowledge of the extent to which the total DNA of a *Mycoplasma* may be duplicated sets. If the DNA content ascribed to *M. arthriditis* (444 million molecular weight units) is, in fact, a single set of genes, and if we consider the average gene as coding a protein of about 500 amino acids, this would allow about 480 genes. Thus, it would appear the *M. arthriditis* may represent a level of complexity about 2 or 3 times greater than our hypothesized hereditary system.

The existence of the *Mycoplasma* doesn't prove the validity of the hereditary system designed here. They help to make it a plausible construction.

Conclusion

There are many alternate assumptions one can make in designing a cell. Any more adaptable cell requires a good many more parts and the imposition of control systems. My point has been to demonstrate that we can design a minimal cell. It contains no magic, self-duplicating, molecules. As a system of molecules with interdependent functions, it has the self-duplicating property we have associated with cells for nearly a century and a half.

With this level of understanding of life processes we are clearly ready to move on to the examination of the kinds of regulatory systems that make possible the more complex organisms. Let us hope that they will prove to be as delightful to understand!

Acknowledgement—The author wishes to thank Ms. Jane K. Glaser for preparing the figures and photograph.

References

DAYHOFF, M.O. 1979. *Atlas of protein sequence and structure, Volume 5, Supplement 3, 1978.* Washington, D.C.: National Biomedical Research Foundation.

HAYFLICK, L. (ed.) 1967. Biology of the mycoplasma. *Annals of the New York Academy of Sciences,* Vol. 142, Article 2. (The articles by Maniloff and Morowitz, and Morowitz, et al. are particularly helpful.)

KENNEDY, D. (ed.) 1965. *The living cell.* San Francisco: W.H. Freeman. (A collection of *Scientific American* reprints).

MOROWITZ, H.J., and TOURTELLOTTE, M.E. 1962. The smallest cells. *Scientific American* 206(3):117.

WATSON, J.D. 1976. *The molecular biology of the gene.* 3rd ed. Menlo Park, California: W.A. Benjamin.