

Molecular Biology and Human Nature

Joshua Lederberg
On the Complexity of DNA

“Molecular biology” is viewed by some as the most arcane of scientific specialties spawned in recent history. At the same time its practical importance for human affairs is recognized today even on Wall Street, and it has made possible an extraordinary coordination and simplification of our views of biology.

To appreciate this requires some patience and self-confidence on the part of the reader. Perhaps even more demanding is the writer’s obligation to strip away non-essential technicalities and jargon to the minimum needed to convey the core advances of recent research. In fact this is possible, as never before, precisely because of the scope of new scientific understanding.

Molecular biology is just biochemistry in modern dress, doing homage to broader aspects of biology that connect closely with chemistry, as well as comprising studies that use the most rigorous of chemical procedures. Thus the molecular biologist is likely to be familiar with the electron microscope, to delve into the intimate structures of cells, to be acquainted with a variety of living organisms in natural history, and to view them in terms of their evolutionary relationships. The traditional specializations of zoology, botany and microbiology are viewed as impediments by molecular biologists, since most of the fundamental principles are common to all life on earth. The chemical substances that are most important in cells are DNA, RNA, and proteins—rather complex aggregates or polymers of simpler units.

“DNA” is so central to our discussion that we can hardly avoid defining it, and explain the acronym “DeoxyriboNucleic Acid”, as also “RNA”, or “RiboNucleic Acid”. These nucleic acids were discovered almost 120 years ago by Friedrich Miescher, as gummy constituents of pus cells. Evidence accumulated that DNA was mainly found in the nuclei of cells (hence the name). RNA was mainly found outside the nucleus, but was chemically related to DNA, containing a certain sugar (ribose) in place of the deoxyribose found in DNA. So a rather intricate history of the unfolding of knowledge is buried in these acronyms. If you are so persuaded, it is

enough to remember that these are important complex substances found in cells, and leave the details to "Scientific American".

Although the occurrence of DNA in nuclei evoked widespread suspicion that it was important for cell biology, its actual function remained mysterious for many years. The importance of DNA was overshadowed in earlier biochemical work by the proteins—the substances found like albumin, hemoglobin, insulin, and many enzymes to be composed of 20 different amino acid building blocks. Many of these, like lysine, tryptophane, methionine, were also found to be of great importance to human nutrition. The human body is unable to synthesize these from simpler foods, and must therefore acquire them directly from plants, or indirectly from meats, milk or eggs. This accent further diverted attention away from nucleic acids, so that they were rather little studied for almost 80 years after Miescher's work. Most biochemists felt that proteins already shown early in the century to be responsible for chemical conversions in the body, i.e., to function as enzymes, would also answer to the remaining secrets of the cell, perhaps, even including those of heredity and of embryonic development.

Except to recall that Gregor Mendel's work, raising the concept of the gene, was contemporary with Miescher, we will leap to more recent developments.

The revolutionary turning point was a paper published February 1, 1944 by a group of scientists from the Rockefeller Institute for Medical Research (Avery with his associates Macleod and McCarty). For 16 years, Avery had been studying a curious phenomenon that had been observed among strains of the pneumococcus (the germ of pneumonia, then a life-and-death threat to health). Extracts of type II, injected into mice together with bacteria of type I, sometimes resulted in live bacteria of type II. In the language of the time, the extract could "transform" the type of germ. Microbiologists and geneticists in those days were highly specialized, and scarcely anyone remarked that this "transformation" might be of great interest to general biology. Avery fully expected the "transforming factor" to be a protein. Instead, it was DNA!

—footnote: Joshua Lederberg, then an undergraduate at Columbia College, was introduced to the 1944 paper by his mentor, the late Professor Francis J. Ryan: it was the cardinal event for the consequent path of his own career. The Rockefeller University, where he is now located, is the present day name of the Rockefeller Institute for Medical Research.—

This finding, together with Avery and others' hints that the "transformation" might be the actual conveyance of gene material from one strain to another, attracted wide interest in the chemistry and biology of DNA, and the recognition that this was the central genetic substance. In today's terminology, we say that genetic information is encoded in the chemical structure of the DNA in the nucleus, in precisely the same sense that a symphony is encoded in the grooves of a recording disk, or that your i.d. is encoded in magnetic material on the back of a credit card or on a computer tape. By 1953, Watson and Crick had worked out many details of the chemical structure of DNA as a "double helix"; and there was no doubt that a new phase of scientific research, molecular biology, was born.

Within the last few years, the biochemistry of DNA has been well enough worked out that it can be used for technological applications. The most important technique is the re-splicing of arbitrary bits of DNA from diverse organisms, the so-called recombinant DNA methodology. This can be used, for example, to manufacture human insulin by implanting a segment (the right segment!) of human DNA into a bacterium, and growing that modified bacterium in a fermenter on a large scale. More important to human welfare, in the long run, will be the further scientific use of these methods for deeper understanding of human biology and disease, in such fields as cancer, developmental disorders and psychiatry.

Finally, the measurement of DNA in human cells (Fig. 1) casts a new philosophical perspective on human nature. Ever since Descartes, biological scientists have been embroiled in arguments over the philosophy of mechanism: that is, whether living phenomena could be understood in principle as the working of a machine, a bit like the clockwork that was evolved during the renaissance. Biologists rarely doubted the leap in complexity from clock to embryo, but their faith that mechanical understanding could eventually be achieved has motivated every constructive step in the progress of biology and medicine.

Before recent DNA research, it was hardly possible to put bounds on just how "complex a machine" the human cell was. Now we can say that it comprises several billion 'characters' worth of information; to date, we can describe less than 0.1% of that data bank. Present technology would require several centuries of effort by all existing laboratories to complete the transcription of that DNA code for a given individual; many times that would be needed to sketch the variety of DNA shown by the diversity of human beings. In principle, all that could be achieved by more and more sophisticated automation, perhaps within the next half-century. The result would be volume upon volume of cryptic text, of which Figure 2. shows just a few leading sentences. (This is a small part of the 16,569 characters of the DNA of the mitochondrion, a small organelle of the human cell.)

As is obvious, this transcript is only the beginning of mechanistic understanding, although the DNA sequence is the actual result of 4 billion years of evolution (yes, roughly one year per character inscribed.) Almost every law of chemistry and physics is involved in the interpretation of that code, its eventual manifestation in the development and bodily structure of the human. This perception of the complexity of human organisms in no way refutes the Cartesian principle. It does indicate how unreasonable is any practical hope of predicting evolutionary outcomes from the first principles of chemistry and physics. That would be far more complex than the practically impossible task of predicting all possible inventions that could be contrived by human intelligence from the same underlying laws. The realistic hope that does inspire and pervade modern science is to understand, in some measure, the marvelous mechanisms that Nature does provide, by reasoning back to how they do work in terms of the properties of proteins, and then in turn to the structure of the DNA embodied in our cells. Just as with history, we can expect reasonable stories to emerge; but we have to be equally skeptical of any claim to rigorously foretelling the future in the face of such huge contingency.

Figure 1.

A PRIMER ON HUMAN DNA

The DNA of each human cell comprises

- 3,000,000,000 units in 2 meters
- 10,000,000 genes possible
(= scope of Encyclopedia Britannica)
- Only about 1% actively code for proteins (rest 'selfish?')
- 100,000 proteins probably make up the constituents of the human body.
- About 1,000 proteins have names AND can be guessed to be present in the body.
- About 100 proteins have been isolated and definitely characterized in human cells.
- If scaled in width to a 1" tape the DNA of a single human cell would stretch round the world.

Figure 2.

An example of the DNA code. The first 200 of the 16,569 characters of human mitochondrial DNA.

```
GATCACAGGTCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCATT
TGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGA
GCCGGAGCACCTATGTCGCAGTATCTGTCTTTGATTCCCTGCCTCATCCTA
TTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGT
```

The alphabet comprises the 4 letters A C G T, standing for the 4 possible base molecules in nucleic acid.