Scientific convergence in the birth of molecular biology

Sunny Y. Auyang

"I myself was forced to call myself a molecular biologist because when inquiring clergymen asked me what I did, I got tired of explaining that I was a mixture of crystallographer, biophysicist, biochemist, and geneticist." Thus explained Francis Crick, who with James Watson discovered in 1953 the double helical structure of DNA, the genetic material[1]

Born of the union of biology, chemistry, and physics, molecular biology is a basic science with intense practical and commercial consciousness. It was barely six decades old on June 26, 2000, when U.S. president Bill Clinton and British prime minister Tony Blair jointly announced its crowning achievement, a draft for the human genome, the "book of life" listing all human genes. "Today we are leaning the language in which God created life," proclaimed Clinton.[2] Others had more mundane concerns. "The genomic gold rush" flashed on the cover of *BusinessWeek*.[3] It was not the first gold rush since the late 1970s, when molecular biology spawned the biotechnology industry that brought human insulin, growth hormone, and other biological drugs and diagnostic tools.[4] Many more are promised. However, the envisioned biomedical revolution is slow in coming, partly because disease phenomena are so complex. To combat the complexity, science is increasingly polling knowledge scattered among various traditional branches.

This chapter examines the scientific contents of molecular biology, highlighting its integrative characteristic. It has roots in biochemistry, genetics, microbiology, and crystallography for molecular structures. As it grows, it draws in other disciplines: medicine, engineering, computer science, and the humanities. The medical sciences reveal important problems and directions of research. Engineering contributes not only equipment and technology but also approaches to address complex functions. Computer science provides methods to analyze torrents of experimental data, interpret results, and manage accumulating information. The humanities participate in addressing ethical, legal, and social issues, which consumed about five percent of the budget for the human genome project.

Convergence and merger of scientific disciplines are not new. The very name of "biochemistry" announces a hybrid. Interdisciplinary research centers already appeared in America after World War II. Even so, molecular biology is outstanding for its scale of and conduciveness to collaboration. The completion of the human genome project marked not an end but a new beginning. Now that the "book of life" is published, scientists face the far more difficult task of reading and making sense of it. For this, molecular biologists are joining with other scientists for integrative studies of life in all levels and perspectives. A 1999 *Science* article on "exploring the

systems of life" reported a building boom of multidisciplinary centers in university campuses, signaling the research trend in the new century: "The convergence of chemistry, physics, biology, and engineering is upon us."[5] It could have added medicine; many of the new centers have pharmaceutical ties.

Unlike most life science disciplines such as physiology and microbiology, molecular biology grew up in the school of science rather than medicine. The initial aloofness enabled scientists to stand back for a broad perspective and to take time to investigate deep and complex problems with no immediate applications. Once basic research begins to explain the mechanisms underlying heredity and other vital processes, molecular biology returns to medicine vigorously and has already transformed the relationship between academic and industrial research.

Stovepipes and missile silos

Among the founders of molecular biology were geneticist George Beadle and biochemist Edward Tatem, who teamed up in the late 1930s to find the relations between genes and enzymes. Tatem's father drew Beadle aside during a visit and expressed concerned about his son: "Here he is, not clearly either biochemist or geneticist. What is his future?" Beadle recalled: "I attempted to assure him – and perhaps myself as well – by emphasizing that biochemical genetics was a coming field with a glowing future." Events proved him right; biochemical genetics has become a pillar of molecular biology. Nevertheless, Beadle noted the difficulty of pulling together disciplines separated by "human limitations and the inflexible organization of our institutions of higher learning."[6]

Although this chapter emphasizes scientific convergence, it should not give the wrong impression that collaborations are easy or unions harmonious. Strives fill the history of molecular biology. The scientists who stood shoulder to shoulder behind President Clinton on that June day in the White House were off camera at each other's throat about the priority in deciphering the human genome. Stories of conceit, greed, and betrayal, common in gold rushes, make best sellers. Instead of repeating them, let us focus on the intellectual and scientific dimensions, which are less sensational but no less important. We would better appreciate the centripetal trend if we realize the strength of the centrifugal forces that are pulling science apart.

As science advances, the richness and complexity of its contents skyrocket. New disciplines and subdisciplines mushroom, each introducing its concepts, jargons, and techniques, and each offering so many intriguing problems as to absorb the whole mind of any single person. Students specialize earlier and earlier in their education, and professionals find it harder and harder to communicate outsider their specialties. As great intellect is not infrequently accompanied by great ego, the justifiable esteem of one's own work sometimes slides into unjustifiable scorn of others' works. Scientific research is everywhere competitive, but with their penchant for candid memoirs and partisan histories, molecular biologists make it seem red in tooth and claw even without the lure of big money. Watson, whose Ph.D. thesis was in genetics, threw darts at almost everyone, not the least geneticists: "You would think that with all their talk about genes they should worry about what they were. Yet almost none of them seemed to take seriously the evidence that genes were made of DNA. This fact was unnecessarily

chemical."[7] Geneticist Gunther Stent described the "tendency of the early molecular geneticists to look down on biochemistry." His own historical account of the origins of molecular biology gave no credit to biochemists and quoted in length negative remarks about them.[8] Biochemist Erwin Chargaff retorted that molecular biologists were merely "practicing biochemistry without a license."[9] Mutual slight hindered progress, but fortunately, it did not prevent molecular biology from emerging as a synthetic discipline.

One characteristic of twentieth century scientific research is its large scale. The number of scientists explodes and research projects become gigantic. The social organizations designed to manage the scientific mass further fragment the structure of scientific knowledge. Research groups, academic departments, corporate laboratories, and government agencies eye each other jealously, each fights to claim more credit, grab more funding, attract more talents, protect or expand its turf, and ensure loyalty of its members. This is reflected in the words "club" and "church" that often pop up in the history of molecular biology. Stent remarked of his own group: "one main characteristic of the members of the American Phage Group was that they didn't believe anything that anyone had said or done before." Historians wondered if such exclusive attitude had delayed the acceptance of DNA as genetic material through the 1940s.[10] Similar concerns are widespread. Many laboratory directors worry about the inclination of groups to dismiss things "not invented here," because such attitude would reduce a laboratory's research productivity, not to mention hindering scientific progress and technological diffusion at large. Administrative reorganization can help to break walls and encourage cultural exchange, but they are not without difficulties. The *Science* article that reported on new multidisciplinary research centers also observed that people who work in cooperative teams may be at a career disadvantage. Their home department insists on them to remember: "Who evaluates you for tenure and the quality of your work?"[11]

Size, specialization, and bureaucracy threaten to turn the scientific community into a bunch of stovepipes each venting its own smoke. Many accounts in science and technology studies turn the stovepipes into missile silos hardened by ideologies: reductionism, vitalism, genetic determinism, cultural determinism, and other -isms more at home in science studies than in science. The ambiguity of their meanings increase in proportion to the animosity they engender. Paradigms, which to most people mean significant conceptions and perspectives, become in science studies incommensurate, islands separated by unbridgeable gulfs. Incommensurability is the most significant feature of scientific paradigms as expounded by philosopher Thomas Kuhn, who regarded allegiants of different paradigms as living in different worlds and paradigm shifts achievable only by something akin to religious conversion.[12] It dovetails into the postmodern doctrine that scientific knowledge is never true but only "true," read true-within-the-culture-in-which-it-is-constructed. Because rational discussions across incommensurate paradigms are deemed impossible, belligerent rhetoric becomes fashionable. Interdisciplinary relations are described not as diffusion or adoption but invasion or conquest.

Fortunately, splintering forces have not prevailed, although they remain strong. Amid all the divergence and diversity, a scientific tendency persists in converging and seeking common understanding. Scientists are and can never be God. They do not have an absolute transcendent position to see the whole world, but can only investigate natural phenomena from their situated human positions. Yet they can try to transcend their particular cultures, communicate, and

acquire broader knowledge. In their attempts to fathom the immensity and complexity of nature, various sciences are like the fabled blind men groping to figure out an elephant. One says it is a moving hose; another, a flapping carpet; a third, a rubbery pillar. If they keep on groping, eventually their hands will touch. And if they believe they share a real world that is much more vast and complex than depicted in their respective culture, if they are not brain washed to insist that their results are incommensurate, they would come to talks, not blows. Communication is not easy. Both sides may have to modify their previous beliefs. Some old timers may quit unreconciled and young minds enter with fresh ideas. Radically new conceptual frameworks may be established to explain the combined phenomena. The integrative process may be long and arduous, sprinkled with bickering and set backs. But it is possible and its partial success fuels scientific growth. Let us examine how convergence led to a most productive science, molecular biology.

Macromolecules at the basis of life

"Molecular biology, then, is the study of how DNA, RNA, and protein are interrelated," summarized David Baltimore in his forward to *Nobel Lectures in Molecular Biology*.[13] In this sense, "molecular biology" refers to a focused science. It is narrower than the casual sense of the phrase meaning the study of life at the molecular level, which applies to most if not all areas in life science. Only the narrow sense is used here. Knowledge gained in molecular biology may contribute to embryology, physiology, and other scientific areas, but these are not part of molecular biology as far as we are concerned.

Organisms come in dazzling varieties, but chemically they are rather homogeneous. Of the millions of chemical compounds available, they are made up of only water, a few hundred kinds of small molecules, and four classes of large macromolecules. A macromolecule is a long chain of linked small molecules, which together consist of thousands to tens of thousands of atoms. Two classes of macromolecules, fats and carbohydrates, serve mainly as foodstuff and building material. They are rather monotonous; their chains, no matter how long, are made up of small molecules arranged repetitiously. The other two classes of macromolecules, proteins and nucleic acids (deoxyribonucleic acid, DNA and ribonucleic acid, RNA), perform heredity and dynamic functions that are the essence of life. Anything but monotonous; their variety underlies the diversity of life forms.

A strand of DNA is a long sequence of subunits, called nucleotide bases, attached consecutively to a backbone made of sugar and phosphate. All nucleotide bases in DNA fall into only four kinds: adenine (A), guanine (G), thymine (T), and cytosine (C). Their irregular order in the sequence constitutes the primary structure of DNA. Two DNA strands running opposite to each other constitute a DNA molecule in the form of a double helix. The bases in the two strands form complementary base pairs: the base A in one strand is always paired to the base T in the other, and G to C. The two members of a base pair bind to each other. The bond in one base pair is rather weak, but millions of base pairs acting together hold the two DNA strands together firmly, not unlike the weak teeth of a zipper acting together to form a strong bind.

DNA accounts for only a fraction of a percent of a cell by weight, but its minute amount belies its vast importance. It performs three vital functions. Its first function, to carry enormous information for specifying heritable traits, is performed by the irregular ordering of its numerous base pairs. Its second function, to serve as a stable template for the almost exact replication of itself, is facilitated by its double helical structure. DNA's third function is to provide detail recipes for the synthesis of proteins. Protein synthesis is a complicated process. Roughly, some DNA base pairs serve as codes that are transcribed into RNA that, after editing if necessary, is translated into a protein or a polypeptide chain.

A protein is made of one or more polypeptide chains. A polypeptide chain is a long sequence of amino acids, each belonging to one of twenty kinds, linked to each other by strong peptide bonds. The linear order of the amino acids in a chain, which constitutes the primary structure of a protein, is specified by DNA. A chain's primary structure determines how it folds up to form higher order structures. These three-dimensional structures have highly specific shapes that are crucial for proteins to perform their specific functions.

If DNA is like the queen bee in a bee colony, then proteins are the worker bees. They are more varied than DNA and more abundant, accounting for more than half the dry weight of most organisms. They constitute the core of life dynamics and perform a host of biological functions. Structural proteins such as elastin endow tissues with strength. Transport proteins such as hemoglobin carry molecules to various parts of the body as needed. Regulatory proteins such as hormones convey signals among cells and maintain them at favorable conditions. Defense proteins such as antibodies recognize and neutralize foreign bodies and pathogens. The most varied and specialized proteins are enzymes, which direct and catalyze almost all chemical reactions in the body. Enzymes are amazingly efficient and can speed up a reaction by ten billion folds. Without them, metabolism and life itself grind to a halt.

DNA resides in each cell of an organism and in the nucleus if the cell has one. The nucleus of a human cell is barely 6 micrometers in diameter, but it contains all the genetic information of a person. The genetic material in a cell, about 2 meters long if stretched out, divides into 46 DNA molecules. Each thread-like DNA molecule coils up to form a chromosome, a sausage-like entity that exhibits a definite pattern of bands when stained with dye. A human has 23 pairs of chromosomes with various sizes. Together the 46 DNA molecules in the human genome contain about 6 billion base pairs. Of these, only 1.5 percent constitutes codes for the structures of proteins and polypeptide chains. Less than 20 percent constitutes regulatory elements that control how and when a code is to be activated for polypeptide synthesis. The remaining DNA is either spacing material between coding regions or "junk" with unknown function.

DNA molecules, chromosomes, base pairs, proteins, and polypeptide chains are clearly physical entities individuated by their spatial separations or physical characteristics. The same cannot be said of genes. To say genes are DNA is saying only that DNA is the stuff genes are made of. It does not answer what counts as *a* gene. To that question no consensual answer exists, but that in no way implies that genes are fictitious constructions. Most criteria for individuating genes now invoke both material and function, and they overlap in essential points. Some definitions designate as a structural gene the DNA regions that contain the codes for a functional polypeptide, and a regulatory gene the DNA regions containing regulatory sites. Other

definitions combine the two and include in a gene all DNA regions, coding and regulatory, that are responsible for the synthesis of a functional product, notably a polypeptide that is a functional part of a protein. Both definitions individuate genes not spatially but functionally. DNA regions of a gene may not be contiguous and DNA regions of two genes may overlap. Because of these and other complexities, gene counting is a difficult task.[14]

When the human genome project began, excited scientists expected to find as many as 100,000 protein-coding genes. They were disappointed by the first draft of the human genome; humans have only 32,000 genes, less than twice the number of genes in the lowly nematode worm. Those who thought it an embarrassment for humans were destined to be further humiliated. Refined counting in the completed version of the human genome found only 20,000 – 25,000 human genes, of which at least 2,000 are pseudo genes that have lost their functions. The small number of human genes relative to human capabilities accentuates the importance of gene regulations, protein dynamics, and developmental complexities.[15]

The number of human proteins is considerably higher than the number of human genes for several reasons. A gene's transcription can be spliced and combined in several ways in the process of protein synthesis. The polypeptide encoded by a gene can function in several proteins. A polypeptide can also acquire small ions and molecules to perform new functions. On top of sheer numbers, the location, timing, and abundance of each type of protein make crucial differences to the welfare of cells and organism itself. All these make proteomics, the study of all proteins and their functions, far more complex than genomics.[16]

We have a glimpse at the intricate relations between DNA and proteins. DNA is the genetic material, but proteins provide the definitions of individual genes. DNA specifies the structures of proteins and the procedures of their synthesis. In return, proteins sensitive to cellular conditions bind to the regulatory sites of DNA, turn it on or off, and control its activities. Some enzymes catalyze DNA replication, others repair damages in DNA, and still others facilitate RNA transcription. Without proteins, DNA is a dead molecule. Without DNA and RNA, proteins cannot be produced. How the intricate dances of proteins, DNA, and RNA drive the dynamic of life is the core topic of molecular biology.

Molecular biology had a long road to travel before reaching the picture of macromolecules presented above. DNA was experimentally isolated in the nineteenth century. Proteins, especially enzymes, had been the staple of biochemistry. However, it was not until the early 1940s that the genetic role of DNA was identified by Oswald Avery's group and the relation of enzymes to genes established by Beadle and Tatem. That was the beginning of molecular biology, although its roots reach back further.

Even at the basic level of macromolecules, life exhibits awesome complexity. Molecular biology approaches genes and proteins from four sides: their proximate and remote functions, their properties and interactions, their microscopic structures, and the processes of manipulating them. For these it integrates and develops, among others, four major disciplines: genetics for functions, biochemistry for properties, crystallography for structures, and biotechnology for processes. The first three disciplines have traditions of their own.

Genetics and gene functions in heredity

Biology, which lacks broad fundamental principles, is more fragmented than physics and chemistry. It has roots in medicine, natural history that emphasizes description and classification, and natural philosophy that emphasizes experimentation, generalization, and underlying mechanisms. Ironically, it was the tradition of natural history that produced the most important arching biological theory, Charles Darwin's theory of evolution by natural selection. It provides the philosophical underpinning for molecular biology. The evolutionary concept of descent by modification from a common ancestor helps to justify several practices in molecular biology. One practice is to investigate simple organisms such as microbes and judiciously extrapolating results to higher organisms. Another is to draw inference from comparing genes and proteins from various organisms. Conversely, the discovery that all organisms share similar molecular structures in their genetic materials lends empirical support to the theory of evolution.

Evolution is possible only if parents are able to pass down some of their traits to their descendents. Most people thought that inheritance of traits is not magical but brought about by some material means. What means exactly? Hippocrates and Aristotle and others had speculated about it, and Darwin added his own conjecture, all to no avail. It was difficult to find the mechanism underlying heredity when the phenomenon of heredity – who inherit what from whom – was itself vague and unclear. Transmission of heritable traits seemed to be erratic, which made it difficult to explain. Were there patterns or rules in the transmission of traits through the complexity of sexual reproduction? If so, then the patterns would provide clues for the underlying mechanisms. Classical genetics started by delineating *patterns* of transmission, molecular biology ended by finding the *mechanism* of heredity.[17]

People study hereditary phenomena in various ways. Practical breeders hybridize plants and animals, aiming to produce particular breeds but not general solutions. Natural historians such as Darwin amass huge amount of miscellaneous data, which are difficult to analyze and extract answers. Experimental biologists design controlled experiments, akin to posing specific questions to nature.

In 1856, three years before the publication of Darwin's *Origin of Species*, Austrian monk George Mendel started a series of simple breeding experiments with a manageable number of variables, so that regularities were easy to discern and experimental results quantified. He crossed peas with easily recognizable traits and counted the number of progenies that expressed each trait. To interpret the data, he proposed a model in which each trait was controlled by a pair of *Merkmal* (character), now called *genes*, one gene derived from each parent. Mendel's results lied dormant until 1900, when it was rediscovered independently by four scientists who were impressed by his way of modeling and data analysis. His model came to be known as Mendel's laws, which launched the science of genetics.

The history of genetics in the twentieth century exhibits two intertwining trends. The first, the miniaturization of experimental organisms, draws in microbiology. The second, what Baltimore called "the materialization of the gene,"[18] draws in biochemistry and crystallography. Classical genetics, which addresses big pictures of transmission patterns, takes a top down view of

heredity. Biochemistry and crystallography, which address molecules and their structures, take an atomistic and bottom up view. Molecular biology bridges the two views. It is facilitated by the study of bacteria and viruses, in which the gap between top and bottom is narrower.

During the late nineteenth century, cytologists put cell fertilization and cell division under the microscope. They discovered that inheritance is transmitted by chromosomes, certain segments of which shuffle and recombine during sexual reproduction. The fruit fly *Drosophila* had huge chromosomes that were easily analyzed with available equipment. It became the choice experimental animal for Thomas Morgan's genetic research program commencing 1910 in the United States.

Morgan's group discovered complex patterns in which some traits linked to each other with various strengths and transmitted as units; for instance, the fruit fly's eye color linked to its sex in inheritance. They systematically compared the patterns of trait inheritance with the shuffling of chromosomes. This method enabled them to associate the genes for various traits with various positions on specific chromosomes, and to associate the linkage between two transmitted traits with the distance between their respective genes on the chromosome. The fruit fly research was an influential forerunner of molecular biology. In anachronistic terminology, it first demonstrated that "genetic information" has a linear physical arrangement that can be mapped. It also produced the first genetic maps. Although crude, chromosome positioning still works in modern genetic mapping. In the human genome project, it provides a framework that anchors detail sequence information.

Locations on chromosomes indicated a cellular base for genes. The materialization of genes took another step forward in 1927, when Morgan's student Herman Muller demonstrated that X-ray caused fruit flies to mutate. Muller did not know how X-ray and other agents affected genes. The knowledge gap between gene functions and mechanisms remained, and the gene in genetic theories was still an abstract entity. Nevertheless, he was convinced that genes were material and foresaw the nature of research that would fathom their secrets: "Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so."[19] His multidisciplinary hope was answered by molecular biologists.

Microbiology and biochemical genetics

Microbiology had advanced much since its nineteenth century genesis discussed in the preceding chapter. Its convergence with biochemistry ensued with the research into the genetics of microorganisms. In 1928, British physician Fred Griffith studied two strands of pneumococci, one not pathogenic while the other caused pneumonia. He discovered that the pathogenic strand, when heat killed, caused the nonpathogenic strand to transform into the pathogenic one. What in the dead bacteria caused living bacteria to transform? In their search for an answer, Americans Oswald Avery, Colin MacLeod, and Maclyn McCarty identified DNA as the genetic material. We will return to this groundbreaking work shortly.[20]

Beadle and Tatem used microorganisms in their experiments that led to the one-gene-one-enzyme hypothesis. In 1946 Tatem's student Joshua Lederberg provided the first demonstration

that bacteria exchange genes by sexual reproduction and other means, thus founding microbial genetics. Genetic thinking enriched microbiology, illuminating important phenomena such as how bacteria evolve resistance to antibiotics. Conversely, microbial experiments opened vast areas for genetics, not the least because microorganisms are more conducive than peas or flies to biochemical analysis. Beadle, Tatem, and Lederberg brought together biochemistry, genetics, and microbiology, three strong voices in the fugue of molecular biology. They shared the 1958 Nobel Prize, the first for molecular biology, if we regard the classical genetics of Morgan and Muller as prehistory that does not address molecules.[21]

Bacteria are single-cell organisms. Even simpler are viruses, which are not cells but merely genes wrapped in a coat of proteins. To reproduce, a virus must get its genetic material into a suitable cell to steal its machineries. The crystallization of the tobacco mosaic virus by Wendell Stanley in 1935 was a popular sensation, because it put life into a form quintessential of inanimate things. For biology, it opened the door to powerful tools in physics such as X-ray crystallography, which eventually revealed the double helix of DNA.

Phages, viruses that infect bacteria, are especially convenient for genetic research. In the 1940s, physicist-turned-biologist Max Delbrück joined microbiologists Alfred Hershey and physician-with-excursion-into-physics Salvador Luria to study phage genetics. The Phage Group organized annual summer schools at Cold Spring Harbor, where phage researchers exchanged scientific ideas, forged social bonds, and recruit talents.[22]

From peas to phages, genetic experiments proceeded toward smaller and simpler organisms. One advantage of the trend is obvious. Peas complete one reproductive cycle in several months, fruit flies in several weeks, and phages in half an hour. The quick turnover of microorganisms enables researchers to select rare strains, collect more data, and design experiments that are more complex.

Using simpler organisms also enables researchers to zero in on important genetic factors unencumbered by overwhelming details. To cut through intractable details and recognize simple cruxes is more common in the physical sciences than biology. Sure, simple solutions are often approximate, but they enable scientists to make the first cut and proceed stepwise to refine approximations instead of being stuck forever in the tar pit of complexity. Beadle found that the relatively simple relation between genes and enzyme production met with considerable resistance from biologists used to thinking about remote gene functions such as milk production of dairy cows. Referring to the "persistent feeling that any simple concept in biology must be wrong," he cited many examples from molecular biology offered the lesson: "Do not discard a hypothesis just because it is simple – it might be right."[23]

Microorganisms live under all kinds of harsh conditions, from hard rocks to undersea volcanoes, which would kill any higher organism. The chemical tricks they evolved to cope with varied environments have become treasure troves for science. Microbes are sources for most antibiotics. They also provide most enzymes that make possible genetic engineering and modern biotechnology, as we will see. Knowledge acquired and experimental techniques developed in microbiology have become indispensable to molecular biological laboratories. Bacteria and viruses constitute an experimental platform, serving as laboratories, factories, vehicles, and tools

for analyzing, manufacturing, and transporting DNA and proteins. They enable biologists to insert a selected piece of genetic material into a cell and study how it works or what effects it has. Such methods led to many important discoveries. In sum, microbiology is now an essential part of molecular biology, even in the study of higher organisms.

The gap between substantive genes and functional genes

The experiment that many molecular biologists cite as the beginning of their discipline was published in 1944 by microbiologists with medical backgrounds: Avery, MacLeod, and McCarty. Baltimore and Watson were not alone in regretting that Avery did not live long enough to be honored by a Nobel Prize.[24] They deserved it; their experiment on the material that effected bacteria transformation provided the first correct and convincing substantive answer on the nature of genes: Genes are DNA.

Deoxyribonucleic acid was isolated in 1869 by Swiss chemist Friedrich Miescher, who later demonstrated that it exists only in chromosomes, the site of hereditary material that cytologists identified. By the 1930s, DNA was known to be a large molecule in the form of a long chain of nucleotides. Beyond that, its structures and functions were poorly understood. Without firm experimental evidence, people believed that the nucleotides repeated themselves regularly, so that DNA was, in the word of Delbrück, "stupid" and incapable of transmitting the infinite variety of traits in heredity. They regarded protein to be the promising candidate for genetic material, because they were familiar with its complexity and variety, although not structures. Into this scientific landscape Avery's experimental result dropped, in the word of Lederberg and Watson, a "bombshell."[25]

Like other discoveries worthy of the name of scientific revolution, Avery's result stirred controversies. Defenders of the received view argued that the DNA in his experiments was contaminated by a trace amount of proteins. Avery answered. Questions remained, but to researchers, new avenues of research were as exciting as final words. The ability to discern significant questions is a hallmark of good scientists. Lederberg, then nineteen and fired by the bombshell, observed that Avery's result opened two important avenues. It indicated a way to investigate the structures of genetic materials, and it called into question the mechanisms by which bacteria exchange genes.[26] He himself pursed the second avenue, which led to his seminal work on bacterial sex. Others, alerted to the importance of DNA, took to analyze it. Avery's result profoundly impressed Chargaff, who switched his laboratory to study nucleic acid.[27] Crick knew about Avery's result and Watson attended an undergraduate course in which it was discussed. Although not certain, the two young men were persuaded enough to put their bets on DNA.[28]

To most classical geneticists, the significance of Avery's experiment was not apparent, partly because their own research had nothing to do with genes being DNA or any other material. Avery's experiment implies a *substantive* conception of genes, which describes genes by their materials, properties, and interactions. Classical geneticists mostly subscribe to a *functional* conception of genes, which defines genes by the roles they play in heredity. A conceptual schism existed in the science of genes, which would only be bridged later by molecular biology.

When you design a home entertainment system, you think about the functions of the players, receivers, amplifiers, and displays, demanding them to produce satisfactory sound and images. You do not care, and probably have no idea, what are in those gadgets. Aside from a few buttons, they are mere black boxes. That was how classical geneticists thought about genes. For Mendel and other geneticists, "gene" meant no more than an algebraic unit in the calculation of trait combinations. They may believe that genes were material, as fruit fly researchers located genes on chromosomes. However, the correlation between genes and chromosome loci was formal and not substantive, with no hint of what it was in chromosome loci that performed the genes' jobs, or how. Aside from their functions of producing patterns of inheritance, genes were black boxes whose substantive contents were beyond the reach or interest of geneticists. This functional conception of genes was enunciated by Morgan in 1934: "What are genes? Now that we locate them in the chromosomes are we justified in regarding them as material units; as chemical bodies of a higher order than molecules? Frankly, these are questions with which the working geneticists has no much concern himself , . . because at the level at which the genetic experiments lie, it does not make the slightest difference whether the gene is hypothetical unit, or whether the gene is a material particle."[29]

Just as I shrug when my friend tells me that his flat panel display is not liquid crystal but plasma, geneticists shrugged when Avery told them that genes are not protein but DNA. Either way, the information would not help their research, so great was the knowledge gap between function and substance. Luria explained why phage geneticists ignored Avery's results: "I don't think we attached great important to whether the gene was protein or nucleic acid. The important thing for us was that the gene had the characteristics that it *had* to have."[30] This was similar to the classical geneticist answer that Morgan gave.

A stovepipe cracked

Phage geneticists were self-proclaimed avant-gardes who disdained classical geneticists for regarding genes as functional black boxes and vowed to open the box and fathom the nature of genes. Yet, when the lid was cracked by Avery, they mostly turned their back to the opening. The dons of the Phage Group knew of Avery's result, but unlike young Lederberg and Watson who deemed it significant, Delbrück said, "you really did not know what to do with it."[31] The chilling response was subject to historical study. Sociologists of science found the major reason in the fact that Avery was an outsider to the Phage Group. The Group applauded the experiment by its members Hersey and Martha Chase, hailing it as the first to identify DNA as the genetic material, although it came years later, provided less evidence, and drew weaker conclusions than Avery's experiment.[32]

Intellectual prejudices abetted social exclusiveness to create stovepipe effects. Stent, a member of the Phage Group, observed that some leaders "were positively hostile to biochemistry." Luria said: "Delbrück and myself, not only were we not thinking biochemically, but we were somehow . . . reacting negatively to biochemistry."[33] The leaders' attitude was especially influential in the culture of what some called the "Phage Church," several members of which reported: "Delbrück deprecated biochemistry, and this influenced some of us to avoid it."[34] When the mechanisms

inside the black box were chemical, rejecting the chemical key led to an impasse for the group. Stent admitted that although the Phage Group drew many correct immediate conclusions from experiments, "the more general and really interesting speculations built upon these first-order conclusions were almost always wrong."[35]

"The impasse was alleviated by the broadening of phage research, sternly governed by Max Delbrück's genius, to embrace a wider range of chemical studies of phage infection," observed Lederberg.[36] The stovepipe began to soften in the early 1950s. One of its consequences was the Hersey-Chase experiment. Another was Luria's sending Watson to learn biochemistry, in Germany. Watson wrote about his doctorate research under Luria: "Initially I had hoped to show that viral death was caused by damage to phage DNA. Reluctantly, however, I eventually had to concede that my experimental approach could never give unambiguous answers at the chemical level. I could draw only biological conclusions. I realized that the deep answers the Phage Group was seeking could be arrived at only through advanced chemistry."[37]

Abroad as a postdoctoral, Watson promptly learned that many roads led to Rome, more than one of which was ignored by the Phage Church. Instead of biochemistry, he was fascinated by Maurice Wilkin's crystallography research on DNA. He made his own way to England and met Crick. Before picking up his story, however, let us pause to trace the chemical origins of molecular biology.

Biochemistry and the materialization of the gene

The functional and substantive concepts of genes may appear to be incommensurate, but they are not. They represent two aspects of the same phenomenon, akin to the trunk and tail of an elephant. Because the phenomenon is so enormous and complex, the two aspects originally seemed disjoint to researchers, who were like blind men groping about the elephant. Yet as Muller realized long ago, a single perspective was insufficient for understanding, and researchers must reach out to other perspectives and other expertise.

The bridging and integration of the functional and substantive gene concepts highlighted "the materialization of the gene." It proceeded both ways. Substantively, the properties and interactions of DNA were elucidated. From the functional side, the remote gene functions considered by classical geneticists were supplemented by proximate gene functions involving the other macromolecule central to molecular biology, protein. On both sides of bridge building, leadership fell on chemistry, biochemistry, and structural chemistry.

Most scientists reject vitalism, but biochemists offer the strongest evidence for refutation. They show by experiments that it is not by mysterious vital forces but by chemical processes that organisms extract energy from the environment to keep alive, move, grow, and reproduce, and these same chemical processes can also proceed in test tubes. Their success in studying biochemical processes in cell-free media disproves any suggestion of vital forces that exit only in cells.

Biochemistry overlaps with molecular biology but goes beyond it to study essential processes in the growth and maintenance of life. It is concerned with the constituents, constitutions, interactions, and functions of all molecules in cells and organisms: carbohydrates, fats, a slew of small organic and inorganic molecules, and of course proteins and nucleic acids. It has long roots in organic chemistry, medicine, and physiology. By the early twentieth century, biochemists had unraveled many chemical processes that break food into smaller and smaller molecules, extracting energy in each step; processes that store and transport the energy to appropriate sites; processes that use available energy to combine small molecules and synthesize large molecules for body tissues, enzymes, and other purposes. Several biochemical processes, each catalyzed by a specific enzyme, succeed each other to form a metabolic pathway. Numerous metabolic pathways drive the dynamics of life.[38]

In 1909, British physician and chemical pathologist Archibald Garrod studied alkaptonuria, in which the patient's urine turned black on exposure to air. Biochemical analysis led him to diagnose the disease's cause as a blockage in a particular reaction in a metabolic pathway. That reaction was catalyzed by a specific enzyme, which healthy bodies produced but patients could not. Geneticists, who tracked the occurrences of the disease in family members, found it to be heritable. Combining genetic and biochemical results, Gerrod concluded that abnormalities in enzymes are heritable.[39]

Are normal enzymes heritable? Beadle and Tatum asked. They induced various mutations in bread mold by exposing them to controlled dosages of X-ray. By culturing the mutants in various nutrient mixtures, they found that in most cases, a strand of mutants lost the normal ability to synthesize a particular nutrient. They argued that X-ray damaged the gene responsible for producing the enzyme that normally synthesized the nutrient. Mutants that lacked different enzymes for synthesizing different nutrients had different genes damaged. Combing their data with genetic knowledge derived from research on fruit fly, they advanced in 1941 the one-gene-one-enzyme hypothesis.[40]

The hypothesis has become one-gene-one-protein and then one-gene-one-polypeptide. Nevertheless, the gist remains that genes are individuated by their chemical and molecular functions. Its enunciation in 1941 introduced a major turn from classical genetics, which mainly studied morphological and behavioral traits such as color of eyes or reproduction of phages. These traits express the physiological functions of genes, which are remote to gene actions. Morgan remarked: "Between the characters that are used by the geneticists and the genes that his theory postulates lies the whole field of embryonic development."[41] Therein lay the wide gap between functional and substantive gene concepts. Biochemical genetics narrowed the gap by sidestepping developmental complexities and going straightly to proteins and enzymes in the most basic processes of growth. Protein synthesis is the proximate function of genes. As genes and proteins interact on the same molecular level, they are susceptible to direct biochemical analysis. This would be the wedge by which molecular biology drove its breakthrough into the secret of life.

To close the functional and substantive gene concepts required knowledge about the physical nature of genes. Biochemists, unlike classical geneticists, appreciated the significance of Avery's identification of DNA as genetic material. To study DNA's properties, Austrian-

American Erwin Chargaff developed highly sensitive techniques to separate the nucleotides. His experiments produced two important results. First, the four kinds of nucleotide occur with varying proportions in DNA, so that they cannot be simply repetitious. Instead of being monotonous, DNA can have enormous variety; the nucleotides succeeding each other in arbitrary order to form a long chain can produce a huge number of possible combinations. This explains DNA's ability to convey huge genetic diversity. Contrary to Delbrück, DNA is not stupid, and a biochemist proved it.

Chargaff's second result is that some regularity does exist in DNA: The percentage of A always equals that of T, and the percentage of G equals that of C. During a visit, Chargaff told Crick about his discoveries. Crick recalled the conversation, "the effect was electric. That is why I remember it. I suddenly thought: 'Why, my God, if you have complementary paring, you are bound to get a one to one ratio.'"[42] Watson's story was a little different. He recalled telling Crick about it before Chargaff's visit, but even after the visit, "there was still a nagging feeling in Francis' mind that Chargaff's rules were a real key."[43] Whatever the details, in gist Crick and Watson both acknowledged that the nucleotide proportions occur in *pairs* provided a strong clue for the *double* helix.[44]

Crystallography and high-order molecular structures

The primary structures of DNA or proteins are long chains of subunits. The chains further form higher-order structures that are functionally important. These structures were mostly unknown in the 1920s, but two powerful approaches were developing. On the theoretical side, quantum mechanics, introduced by German physicists Erwin Schrödinger and Werner Heisenberg, became sophisticated enough to illuminate complicated molecular structures. Although unable to make detailed predictions, it provided knowledge on the strengths, distances, and directions of various kinds of chemical bonds between atoms in a molecule. The theoretical insight, aided by balls and sticks and other mock-ups, enabled scientists to build models for representing and speculating about molecular structures. On the experimental side, X-diffraction, pioneered by British physicists William and Lawrence Bragg, were being applied to biological materials after they were available in crystal forms. Using the interference of light reflected from various crystal plans, X-ray crystallography produced diffraction pictures, from which one could compute the three dimensional spatial arrangements of molecules in the crystal. At that time, X-ray resolution and computation power were rather crude for tackling the complexity of biological molecules. Crystallographers relied on model building to interpret their data.[45]

The British, on the tradition established by the Braggs, dominated the structural research on biological molecules. William Astbury did some measurements on DNA in the late 1920s. However, the major focus of molecular structuralists, including Asbury, was on proteins, because structural conformity of proteins had important functional ramifications. Maurice Wilkins, later joined by Rosalind Franklin, were almost loners devoted to DNA before Crick and Watson burst on the scene.[46]

The only group able to compete with the British was that led by American physical chemist Linus Pauling in Caltech. All researchers relied on both experiment and theory. Pauling, with

his profuse imagination and profound understand of chemical bonds, was more inclined to use simple physical arguments to establish structural constrains on model building. In 1951, he identified an important secondary structure of proteins as a helix.[47]

The Britons who studied the molecular structure of biological materials were the first to call themselves molecular biologists. Many of them had physics backgrounds. This characteristic they shared with many American phage geneticists, who later scrambled to appropriate the name of molecular biology. However, as John Kendrew remarked, "in the early days the two schools were almost entirely isolated from each other."[48] The first major break in the stovepipes was Watson's migration from one to the other.

A few months after Watson arrived in Germany to learn biochemistry, he became fascinated by Wilkins's X-ray picture of DNA. It revealed to him that DNA had regular structures susceptible to scientific study. Primed, he was further excited by Pauling's recent discovery of the helical structure of proteins. He found most of Pauling's reasoning over his head, but he knew what he would have to learn and what road to take. Unable to join Wilkins or Pauling, he went in the fall of 1951 to Cambridge University, which would become a hothouse in molecular biology boasting a dozen Nobel laureates.[49]

Watson joined the crystallography group of Lawrence Bragg, Max Perutz, and John Kendrew. Most members of the group were devoted to protein structures, but he found Crick, who was also interested in DNA.[50] Eleven years later the group would have a reunion in Stockholm, where Kendrew and Perutz received the Nobel Prize in chemistry and Crick, Watson, and Wilkins the Prize in Physiology or Medicine. The Nobel Committee seemed to indicate that molecular structures of proteins and DNA are equally important.

Shortly after the two young men met, they decided to, in Watson's words, "imitate Linus Pauling and beat him at his own game." Watson described in his autobiography how Crick tirelessly taught him crystallography and physical chemistry. He hardly mentioned any contribution of his phage genetic knowledge in discovering the double helix. Instead, he imagined his chief rival to be Pauling, mastery and application of whose insights in model building would be his contribution to the work.[51]

Crick and Watson did not perform their own X-ray experiments but rely on the data of Wilkins and Franklin. While Crick and Watson collaborated in hearts and minds, Wilkins and Franklin were stuck in personal antagonism and a turf war over the exclusive priority to work on DNA. Frictions, competitions, setbacks, and breakthroughs, complicated by gender, made the story of the two pairs one of the most told story in the history of science.

The double helix model of DNA, published in 1953, provided plausible applications of information, template, and other ideas then floating around. The double helix was easy to depict graphically. Now the gene became neither something abstract nor some slimy chemical but an entity that everyone could visualize. It caught the imagination and has become "the Mona Lisa of modern science," in the words of an art historian. As the symbol of not only molecular biology but also life itself, the double helix now appears everywhere, from science to the arts, from architecture to cinema, and of course advertisements.[52]

Striking as it was, the double helix was initially a little more than a theoretical model. It appeared credible and enjoyed some experimental support, but the support was not very strong. The X-ray data was crude and susceptible to alternative interpretations; Pauling had earlier proposed one. Watson and Crick's model was brilliant, but it had to pass numerous experimental tests to be accepted scientifically. This it did with flying color.

DNA as a stable template for self-replication

The most novel feature of the double helix was not helicity but duplicity, especially bonds between complementary base pairs that hold two DNA strands together like a zipper. When the two strands are "zipped" up in the double helix, they are remarkably stable. When they unzip in cell division, the exposed nucleotide bases in each strand constitute a template that attracts complementary nucleotides to form a new DNA strand. Thus the model provides an intuitive explanation for the stability and replication of genetic material. Crick and Watson went so far as to suggest that DNA replication could proceed spontaneouly all by itself.

Arthur Kornberg, a physician turned biochemist, began in 1954 to synthesize DNA. He found that nucleotides assembled only in the presence of a DNA molecule, which served as a template. Thus the double helix model was essentially correct. However, Crick and Watson had grossly underestimated the complexity of replication mechanisms. DNA replication could not proceed spontaneously but depended on a battery of enzymes executing complicated functions. To fathom them took molecular biologists more than three decades, with Kornberg being a central figure in the research. Together, DNA and enzymes operate as a remarkable copying machine for genes. These biochemical processes provide substantive explanations of how DNA performs its genetic function of transmitting heritable information.[53]

Kornberg began his research independently of the double helix model. He disbelieved the model at first, although his experimental results changed his mind. More than one approach to DNA existed, and testing theoretical models was not the top priority of most experimental chemists or biologists. What motivated Kornberg was the strong tradition in chemistry of synthesizing molecules of increasing complexity, the same tradition that led Spanish biochemist Stevero Ochoa to synthesize RNA and Indian-born biochemist Gobind Khorana to synthesize chains of nucleotides and amino acids. The chemists' technical reasoning depended little on theoretical models beyond DNA being a long chain of nucleotides.[54] Their account of intellectual influence seemed to be accepted by the scientific community. For their works in biochemical synthesis, Kornberg and Ochoa won the Nobel Prize for Physiology or Medicine in 1959, three years before Crick, Watson, and Wilkins got theirs for the double helix, and ten years before Delbrück, Hersey, and Luria got theirs for phage research.

"The biochemical phase of genetics," as Kendrew called it, did not end with the discovery of the double helix.[55] Biochemistry assumed increasingly heavy roles in discovering the mechanisms for DNA replication, RNA functions, protein synthesis, and gene regulation, works that constitute the core of molecular biology.

DNA and protein synthesis

How does DNA specify the structures of proteins? How are proteins synthesized under genetic instructions? How is synthesis regulated so that proteins with the right functions are produced at the right place and the right time to satisfy cellular requirements? Answers to these questions would close the gap between functional and substantive conceptions of genes. The first answers came in 1959 from François Jacob and Jacques Monod. One was a microbial geneticist, the other a microbial biochemist. They teamed up upon realizing that similar general regulatory mechanisms underlay the superficially disparate phenomena they were separately studying. Their collaboration resembled that of Beadle and Tatem, only it occurred in not America but France.

One phenomenon Jacob and Monod investigated was that bacteria produced the enzyme specific for lactose digestion only in the presence of lactose. This seemed intuitive; enzyme synthesis requires energy, which is wasted when the synthesized enzyme has nothing to digest. But how do bacteria manage to "economize?" How does the presence of lactose induce bacteria to start synthesizing the enzyme required to digest it? By introducing mutations into the bacteria, Jacob and Monod found that the microbial ability to respond to environmental changes is genetic. Mutant bacteria that lose this responsive ability persistently produce the enzyme for digesting lactose, even when it is not available. These and many other experimental results led them to discover regulatory genes.

Genes have not only *structural* but also *regulatory* functions. Some genes do specify the structures of proteins. Other genes, called operators, act as "switches" controlling the activity of the structural genes. The operators are switched on or off by special proteins sensitive to environmental conditions, thus regulating the production of proteins. The gene for lactose enzyme, for instance, is switched on by a special protein only in the presence of lactose. Jacob and Monod discovered several regulatory patterns. With the help of protein structure crystallography, they explained a mechanism by which regulatory proteins interact differently with DNA under different cellular conditions. For their discovery of genetic regulation, Jacob, Monod, and Lwoff won the Nobel Prize in 1965.[56]

Research on the regulation of protein production also shed light on the biochemical processes of protein synthesis. At that time, molecular biologists knew that although DNA must be involved in protein synthesis, its involvement cannot be direct. DNA resides in the chromosomes inside the cell nucleus, but protein synthesis occurs outside the nucleus at what are now called ribosomes, which are rich in RNA. In 1953, Alexander Dounce suggested that DNA serves as a template for RNA, which in turn becomes the template for proteins. It anticipated the "central dogma" in molecular biology asserting the flow of genetic information from DNA via RNA to protein.[57]

Dounce's hypothesis was intuitive, except for a hitch. The structure of RNA as known at that time made it seem unable to carry the required information. Like DNA, RNA is a chain molecule with four nucleotide bases attached to a sugar-phosphate backbone. Its then-known structure differed from DNA in four ways, three of which did not disable it from carrying the

required information from DNA to protein, the fourth did. The first innocuous difference is that RNA's backbone has a different sugar. The second difference is in one of its bases; RNA has uracil (U) in place of DNA's thymine (T). U and T share the same base-pairing properties. If RNA is synthesized on a DNA template, its base ordering preserves DNA's base ordering. The third difference from DNA is that RNA does not normally form double helices. Besides these innocuous features, all RNA molecules known then shared a fourth characteristic that made them unlikely to be carriers of genetic information. They were either too small or too "stupid" because they contained equal proportions of the four nucleotide bases. We now know that the abundant stupid RNA at the ribosome serves as the "reading head" for translating the genetic information from DNA into proteins. Yet it is unable to carry information. Does another kind of RNA, larger and more various, exist for the job?

Jacob and Monod's research answered affirmatively. A messenger RNA, as it is called, exists but is ephemeral. When a gene switches on, it serves as the template for assembling nucleotides into a messenger RNA, which bears a transcription of DNA's genetic code. The messenger RNA moves to the ribosome, where its code transcription provides the recipe for synthesizing a new polypeptide chain. After it accomplishes its mission, the messenger RNA disintegrates; if it persists, protein synthesis would not stop when the switch at the DNA turns off.

The messenger RNA escaped detection for a long time because of its transience. Once scientists suspected its existence for a particular function, however, they quickly designed experiments to catch it at its job. By 1960, many experiments firmly established the route of information flow from DNA through messenger RNA to protein. The route holds for almost all organisms, although in higher organisms messenger RNA undergoes significant splicing and editing before being translated into protein.[58]

Genetic information and code

Literate people had long been familiar with the idea that irregular permutations of a small number of signs can carry enormous information. Permutations of 26 alphabets, for instance, generate endless tests in English. The notions of genetic information and code were adumbrated in Schrödinger's 1944 book, *What Is Life*, the influence of which was acknowledge by Cricks, Jacob, Watson, Wilkins and other pioneers of molecular biology. Schrödinger speculated that the aperiodic arrangements of atoms in chromosomes contained some kind of "code-script" for the organism. It did not take too many atoms to code for enormous organismic varieties, he argued, pointing to the Morse Code that by permutations of dots and dashes generated unlimited messages.[59]

A telegraphic message is a linear string of dots and dashes. Similarly, genetic information resides in the one-dimensional primary structures of DNA, not its secondary double helix structure. A double helical DNA could be stupid, its four nucleotides repeating themselves regularly. However, several experiments had refuted DNA's stupid image before the discovery of its double helical structure. The first experiment was the identification of DNA as the genetic material. More directly, Chargaff's data demonstrated that the nucleotides occurred in irregular proportions.

Much insight also came from protein research. After a decade of labor, British chemist Frederick Sanger had discovered the exact sequence of amino acids in the protein insulin and showed that the irregular ordering of amino acids could not be determined by general principles. If, according to the one-gene-one-protein hypothesis, the sequence of amino acids in a protein somehow came from the sequence of nucleotides in a DNA, then Sanger's result would imply irregular ordering of nucleotides. Sanger was, like Pauling, among the few who received two Nobel Prizes. He got one for sequencing protein and later another for the more difficult task of sequencing DNA. The DNA sequencing techniques he developed would become the workhorse in the Human Genome Project.

Crick, who was familiar with Sanger's work on protein sequence, suggested with Watson in 1953, "the precise sequence of the bases is the code which carries the genetic information."[60] They were vague in their first allusion to a genetic code. Soon a surprising letter came from theoretical physicist George Gamow, who stated and proposed an (unsatisfactory) solution for what Crick later called the "coding problem." The problem asked how the sequence of four nucleotide bases on a stretch of DNA formally specifies the sequences of twenty amino acids on the protein to be synthesized. It is analogous to asking how a string of dashes and dots in a Morse code specifies a string of English words.

The genetic code was deemed one of the most important scientific problems in the twentieth century. It was suspected, and later confirmed, that the code was universal and shared by almost all life on earth. Its deciphering would give hint on who we are and how we came about. Not surprisingly it became a hot research topic. Molecular biologists attracted to it followed two major approaches, which can be called the dry and the wet.

The dry approach, centered around Crick, used mainly theoretical methods, aided by general constraints based on known protein structures, to frame clever models. It was embraced by the expanded Phage Group, which called itself "the Information School."[61] Many elegant codes were proposed, admired, and rejected. A few general results were proved. One asserts that each amino acid is coded by a triplet of three nucleotide bases, the minimum number required. Because the permutation of 4 bases makes 64 distinct triplet code words and proteins are made up of only 20 amino acids, the genetic code contains more than one code word for each amino acid. What the code words are, theorists were unable to determine.

The information school attracted much publicity. Its effort on the coding problem was, after Watson and Crick's modeling of the double helix, most responsible for the cult of theoretical superiority in molecular biology. Yet information theorists generated more heat than light. The scientific feat of breaking the genetic code was accomplished not by them but by those whom they slighted, biochemists pursing the wet approach of experimentation. After the dust settled, Crick, whom Watson had never seen in a modest mood, shared his lesson in Cold Spring Harbor Symposium: "One of the reasons that I enumerated . . . the early history of the code was to show how little theory was able to contribute."[62]

The stovepipe missed again

If information theorists were like mathematicians who tried to decipher an enemy code from tidbits of intercepted messages, then biochemists were like special agents who tried to capture the cipher machine itself. They had worked on protein synthesis for a long time and identified many relevant cellular ingredients and processes, some of which would turn out to be physical conveyer of the genetic code.

In 1959, American biochemist Marshall Nirenberg made what his friends called a "suicidal" decision: to work on a problem dominated by large research groups in prestigious institutions, the coding problem. He and a German visitor Heinrich Matthaei shared a small laboratory and a large passion. They decided to use newly developed biochemical procedures for studying bacteria extracts in cell-free media and observe what would happen when they add various kinds of RNA that may carry genetic information. After two years of struggle, they struck gold. A monotonous protein consisting of only one kind of amino acid was synthesized in the presence of a RNA consisting of only one kind of nucleotide. They had deciphered the first code word: the base triplet UUU codes for the amino acid phenylalanine. More important than the first crack on the genetic code, their clever experimental design pointed out a promising way to decipher the other 63 code words: Add RNA consisting of a particular base triplet and see that protein is synthesized.[63]

Nirenberg's application to the 1961 Cold Spring Harbor Symposium was rejected. He went to an international conference in Moscow and spoke to an almost empty tiny room. "He did not come out of the phage group – just had no connections with those people," a molecular biologist explained to Horace Judson, whose *Eighth Day of Creation* is a classic on the history of molecular biology. It was almost twenty years apart, Judson mused, but Nirenberg's reception was reminiscent of Avery's when he told them that genes are DNA. Even years later, neither found any place in the history of molecular biology as narrated by a prominent writer of the Phage Group.[64] Other historians are fair. Judson gave Nirenberg and Matthaei ample coverage in his book and chose for chapter title the remark of another molecular biologist: "He wasn't a member of the club."[65]

Fortunately, science is not a mere social construction of an exclusive club; scientific contents do count. The papers of Mattaei and Nirenberg would be published, and before that, scientific objectivity asserted itself. One of the few who heard Nirenberg in Moscow, feeling "blowed over by it," told Crick. Crick had never heard of these two young no-names, but after talking to Nirenberg, invited him to speak again for the conference. This round the hall was packed and galvanized. Immediately a race ensued to decipher the remaining 63 codons. It was easier said than done. UUU was the easiest of the codons; other triplet combinations of nucleotides were far more difficult to obtain chemically. Khorana and Ochoa adapted their expertise to synthesize RNA with specific base sequences. Nirenberg devised ingenious ways to ascertain accurately the base sequence of available RNA. Despite wide mobilization, it took five more years to crack the entire genetic code. For their achievements, Nirenberg, Khorana, and Holley won the Nobel Prize in 1967, one year before the venerable leaders of the Phage Group got theirs.

Genetic engineering and biotechnology

Chemistry, genetics, and microbiology all have stronger inclinations towards applications than physics. Although molecular biology concentrated on basic science during its first two decades, the practical tradition of its roots showed up in Tatem's 1958 Nobel lecture. "With a more complete understanding of the functioning and regulation of gene activity in development and differentiation, these processes may be more efficiently controlled and regulated, not only to avoid structural or metabolic errors in the developing organisms but also to produce better organisms. . . . This may permit the improvement of all living organisms by processes which we might call biological engineering."[66]

Molecular biology, the science of genetic materials, spawned genetic engineering and biotechnology. It is akin to materials science and engineering, which also rose to prominence after World War II and spawned a cutting-edge technology, nanotechnology. Biotechnology and nanotechnology, both science intensive and manipulate molecules, reinforce each other. Both are manifestly multidisciplinary. The U.S. National Research Council in 1989 defined materials science and engineering as the strong interrelationship among material "properties, structure and composition, synthesis and processing, and performance."[67] The four aspects are present equally in molecular biology.

We have discussed how molecular biology addresses the properties, structures, and functions of proteins and nucleic acids. Instruments and processing techniques, although sometimes left out of the core of scientific content, have been indispensable to science ever since Galileo turned his telescope toward the heavens and Hooke peered down the microscope at cells. Molecular biology is no exception. Since its inception, it has depended on ultracentrifuge, electrophoresis, X-ray diffraction, and electron microscope. Most of these instruments, which make essential use of physical properties such as mass and electric charge, are applicable to many materials besides the biological.

As molecular biologists acquired more knowledge about the properties and interactions of biological macromolecules, they began to use these as tools peculiar to their topics of investigation. Synthesis, manipulation, and processing of macromolecules, the fourth aspect of molecular biology, accelerated scientific progress. In the 1970s, it engendered genetic engineering, which aims to turn natural functions of macromolecules into desired performances for serving human needs and wants. A biotechnology industry grew, marked by social controversies and volatile stock markets. Industrial-styled scientific research emerged, exemplified by the Human Genome Project. The rank of molecular biologists swelled almost a hundred fold. Sensational coverage in the popular media, however, missed much of the basic science involved, which boasted a slew of Nobel Prizes.[68]

Central to genetic engineering, also called recombinant DNA technology, is the abilities to design and manufacture specific DNA segments and insert them into living cells to perform certain desired functions. These abilities take decades to develop and refine, as new scientific discoveries lead to new technological inventions. They are mainly biochemical treatments informed by the requirements of gene functions, especially the function of producing proteins that perform desired physiological functions.

Manipulation of DNA began in the 1950s with Kornberg's attempt to synthesize it. The research led to the discovery of several enzymes that work on DNA. One enzyme, DNA polymerase, assembles nucleotides on an exposed strand of DNA to form a complementary strand. Another, ligase, fuses the backbones of two DNA strands to form an integral DNA. More enzymes were discovered in two other Nobel-prized works. Bacteria yield enzymes that can cut DNA molecules at specific nucleotide sequences. They are called restriction enzymes because they restrict the growth of viruses in host bacteria by cutting up viral DNA. Many restriction enzymes leave exposed nucleotides at the cut ends, which can stick DNA fragments with complementary ending sequences. Hundreds of restriction enzymes enable scientists to cut DNA at desirable places and paste together selected pieces. Retrovirus, a kind virus of which HIV is the most well known, yields another potent chemical tool. Unlike other organisms, retroviruses carry their genetic information not in DNA but in RNA. They also produce an enzyme, the reverse transcriptase, which transcribes their RNA into DNA so that their genes can merge into their hosts' DNA for reproduction. Using reverse transcriptase, scientists can start with the messenger RNA of a desired protein and turn it into DNA.[69]

Enzymes enable molecular biologists to cut and paste DNA at will. For practical purposes, the designed DNA has to be mass-produced. Two general methods of multiplying DNA, introduced a decade apart, won Nobel Prizes. The first uses living cells as tiny DNA factories. Pieces of designed DNA are inserted via retrovirus or bacteria genetics into hosts, which can be single-celled bacteria or cultured mammalian cells. As the cells grow and reproduce, they not only replicate the inserted DNA but also produce the protein that DNA coded for. The proteins, harvested from the cell culture, are desirable products. At first, genetic engineering generated much anxiety about the possibility that the engineered microbes could escape from the laboratory and cause biological catastrophe. When proved safe, it got the biotechnology industry off the ground in the late 1970s.

Cellular amplification of DNA met a stiff competitor in 1983. Polymerase chain reaction can amplify trace amount of any DNA fragments in test tubes by using polymerase and other biochemical means. It rapidly found many uses, such as diagnostics kits for genetic diseases or genome sequencing.[70]

Genetic engineering has many applications, from genetically modified crops that prolong the green revolution to DNA fingerprinting that revolutionizes forensics. It has also changed the pace of basic research in molecular biology. Restriction enzymes once kept hundreds of front-line research scientists busy; now they are sold as reagents in bottles. It took years to sequence the first gene, most of time spent in preparing suitable DNA fragments. Now thousands of sequenced DNA fragments are individually identified, arranged in an array, and fixed on a chip, commercial available for experiments. All these provide the technological infrastructure for the science. Despite the acceleration in discoveries it affords, however, the wait for envisioned biomedicine is far from over. Life is far more complex than molecular biologists first anticipated, as we see in the following chapter.

Notes

1. Crick quoted in G. S. Stent. That was the molecular biology that was. *Science*, 160: 390-395 (1968).
2. Quoted in K. Davies. *Cracking the Genome*. New York: The Free Press (2001), p. 236.
3. *BusinessWeek*, June 12, 2000, pp. 148-162.
4. In Genentech's 1980 initial public offering, price per share shot up from $35 to $89 in 20 minutes, breaking the Wall Street record for fastest price rise. Since then biotech stocks had a roller coaster ride. The boom around 1990 was called the "cloning gold rush." See Office of Technology Assessment, *Biotechnology in a Global Economy*, OTA-BA-494 (1991).
5. R. F. Service. Exploring the systems of life. *Science*, 284: 80-83 (1999).
6. G. W. Beadle. Biochemical genetics: some recollections. In *Phage and the Origins of Molecular Biology*, eds. J. Cairns, G. S. Stent, and J. D. Watson. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (1966), pp. 23-32. Beadle quoted in M. Morange. *A History of Molecular Biology*. Cambridge, MA: Harvard University Press (1998), p. 26.
7. J. D. Watson. *The Double Helix*. Norton Critical Edition, ed. By G. S. Stent. New York: Norton. (1980), p. 46.
8. Stent. That was the molecular biology that was.
9. E. Chargaff. *Heraclitean Fire*. New York: Rockefeller University Press (1978), p. 140.
10. F. H. Portugal and J. S. Cohen. *A Century of DNA*. Cambridge: MIT Press (1977), p. 184.
11. Service, Exploring the systems of life.
12. T. Kuhn. *The Structure of Scientific Revolution*. Chicago: University of Chicago Press (1962), pp. 148-152.
13. D. Baltimore, ed. *Nobel Lectures in Molecular Biology 1933-1975*. New York: Elsever (1977), p. viii. The term "molecular biology" was first used by Warren Weaver in 1938 to describe certain programs funded by the Rockefeller Foundation, where it simply meant the application of techniques developed in the physical sciences to investigate life processes, (W. T. Weaver, Molecular biology, origins of the term. *Science* 170: 591-592, 1970). The first scientific practitioner to call his work "molecular biology" was William Astbury, who used it before 1950 to mean the sturdy of structures, functions, and genesis of biological molecules, (W. T. Astbury, Molecular biology or ultrastructural biology? *Nature* 190: 1124, 1961). Both usages had a broader meaning than Baltimore's definition. The broad and narrow senses of "molecular biology" were distinguished and explained by Crick in H. F. Judson. *The Eighth Day of Creation*. New York: Simon and Schuster (1979), p. 203.
14. M. Snyder and M. Gerstein. Defining genes in the genomics era. *Science*, 300: 258-60 (2003).
15. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431: 931-945 (2004).
16. M. Tyers and M. Mann. From genomics to proteomics. *Nature*, 422: 193-197 (2003).
17. A. H. Sturtevant. *A History of Genetics*. New York: Harper & Row (1965).
18. Baltimore, *Nobel Lectures*, p. vii.
19. Muller's 1921 remarked quoted in Judson, *Eighth Day*, p, 49.
20. M. McCarty. *The Transformation Principle*. New York: Norton (1995).
21. M. Singer and P. Berg. George Beadle: from genes to proteins. *Nature Review Genetics*, 5: 949-954 (2004). R. Olby. *The Path to the Double Helix*. Seatle: University of Washington Press (1974), section III.

22. Cairns, Stent, and Watson, *Phage*.
23. Beadle, Biochemical genetics.
24. Baltimore, *Nobel Lecture*, p. viii. J. D. Watson. *DNA, The Secret of Life*. New York: Knopf (2003), p. 39.
25. Portugal and Cohen, *DNA*, ch. 1. J. Lederberg. What the double helix (1953) has meant for basic biomedical science: a personal commentary. *JAMA* 269: 1981-5 (1993). Watson, *DNA*
26. J. Lederberg. Postmature scientific discovery? *Nature*, 324: 629-631 (1986).
27. Judson, *Eighth Day*, p. 40, 94.
28. Judsson, *Eighth Day*, p. 47. Watson, *Double Helix*, pp. 12-13, 18.
29. Morgan, Relation of genetics.
30. Luria in Judson, *Eighth Day*, p. 63.
31. Delbrück quoted in Judson, *Eighth Day*, p. 60.
32. Judson, *Eighth Day*, pp. 57, 627-628. Morange, *Molecular Biology*, pp, 47-49.
33. Luria in Judson, *Eighth Day*, p. 62.
34. Cairns, at al. *Phage*, p,158. See also pp. 22, 148. Watson, *Double Helix*, pp. 18, 79.
35. Stent. That was the molecular biology that was.
36. Lederberg, *JAMA*.
37. Watson, *DNA*, p. 41.
38. J. S. Fruton. The emergence of biochemistry. *Science*, 192: 327-334 (1976).
39. G. W. Beadle. Genes and chemical reactions in Neurospora, in Baltimore, *Nobel Lectures*, pp, 51-63. Morange, *Molecular Biology*, pp. 21-22.
40. M. Singer and P. Berg. George Beadle: from genes to proteins. *Nature Review Genetics*, 5: 949-954 (2004). Beadle. Genes and chemical reactions.
41. T. H. Morgan. The relation of genetics to physiology and medicine. In Baltimore, *Nobel Lectures*, pp. 3-18.
42. Crick quoted in Olby, *Path*, p. 388. See also Judson, *Eighth Day*, p. 143.
43. Watson, *Double Helix*, pp. 75-6, 84.
44. E. Chargaff. Building the Tower of Babble. *Nature*, 248: 776-779 (1974). Olby, *Path*, pp. 211-221. Judson, *Eighth Day*, 94-96.
45. Hess, Origins. Portugal and Cohen, *DNA*, ch. 10.
46. Kendrew, Information and Conformation. Olby, *Path*, ch. 16.
47. J. D. Bernal. The pattern of Linus Pauling's work in relation to molecular biology. In Rich and Davidson, *Structural Chemistry*, pp. 370-379. Olby, *Path*, ch. 17.
48. J. C. Kendrew. How molecular biology began. *Scientific American*, 216(3): 141-144 (1967). Stent observed: "prior to April 25, 1953, no member of the Phage Group thought of or referred to himself as a 'molecular biologist.'" Preface to Watson's *The Double Helix*, p. xi.
49. E. Pennisi. A hothouse of molecular biology. *Science*, 300: 278-282.
50. Watson, *Double Helix*. Morange, *Molecular Biology*, ch. 11.
51. Watson, *Double Helix*, p. 32-34. Watson opposed to mentioning the genetic implications of the double helix in their first paper; it was added at Crick's insistence, (F. Crick. The double helix: a personal view. *Nature*, 248: 766-710, 1974). Their second paper on the genetic implications of DNA's structure made no reference to the genetic literature but relied on intuitive general notions such as template and replication.
52. M. Kemp. The Mona Lisa of modern science. *Nature*, 421: 416-420 (2003).

53. A. Kornberg. *For the Love of Enzymes*. Cambridge, MA: Harvard University Press (1989), p. 273, ch. 5. H. Echols. *Operators and Promoters: The Story of Molecular Biology and Its Creators*. Berkeley: University of California Press (2001), ch. 4.
54. Kornberg, *Enzymes*, pp. 121, 139, 153. Morange, *Molecular Biology*, ch. 20.
55. J. C. Kendrew. Information and conformation in biology. In *Structural Chemistry and Molecular Biology*, eds. A. Rich and N. Davidson. San Francisco: Freeman ('68), p.187-97.
56. Ullmann, A. and Lwoff, A. eds. *Origins of Molecular Biology: A Tribute to Jacque Monod*. New York: Academic Press (1979). Echols, *Operators and Promoter*, ch. 3.
57. Morange. *History*, pp. 140-142, 126-127. The name "central dogma" was introduced by Crick in 1957. However, Crick's version of the dogma stipulates only the one-way information flow from nucleic acids to proteins, not from DNA to RNA, so that it is not invalidated by the discovery of retrovirus where information flows from RNA to DNA. (F. Crick. Central dogma of molecular biology. *Nature*, 227: 561-563, 1970).
58. Morange, *Molecular Biology*, ch. 13. The processes of protein synthesis are much more complicated in eukaryotic cells, where the messenger RNA undergoes complicated editing.
59. E. Schrödinger. *What is Life*. Cambridge: Cambridge University Press (1944), pp. 22, 65.
60. J. Watson and F. Crick. Genetical implications of the structure of deoxyribose nucleic acid. *Nature*, 171: 964-967 (1953).
61. Stent, That was the molecular biology.
62. Crick quoted in Echols, *Operators and Promoters*, p. 14.
63. M. Nirenberg. Historical review: deciphering the genetic code – a personal account. *Trends in Biochemical Sciences*, 29: 46-54 (2004). Judson, *Eighth Day*, pp. 469-489. Echols, *Operators and Promoters*, pp. 21-34.
64. Stent, That was the molecule biology. It has been much criticized, see for example Morange, *Molecular Biology*, p. 132.
65. Judson, *Eighth Day*, pp. 480, 481.
66. E. L. Tatum. A case history in biological research. *Science*, 129: 1711-1715 (1959).
67. Quoted in S. Auyang. *Engineering – an Endless Frontier*. Cambridge: Harvard University Press (2004), p. 243.
68. H. F. Judson. A history of the science and technology behind gene mapping and sequencing. In *The Code of Codes*, eds. D. J. Kelves and L. Hood. Cambridge: Harvard University Press (1992), pp. 37-82.
69. Echols, *Operators and Promoters*. Ch. 10. S. Wright. Recombinant DNA technology and its social transformation, 1972-1982. *Osiris*, 2: 306-360 (1986).
70. Morange, *History*, ch 16. K. B. Mullis. The unusual origin of the polymerase chain reaction. *Scientific American*, 262(4): 56-65 (1990).