Essays of an Information Scientist: Of Nobel Class, Women in Science, Citation Classics and Other Essays, Vol:15, p.367, 1992-93 Current Contents, #39, p.3-12, September 27, 1993

Current Comments®

EUGENE GARFIELD
INSTITUTE FOR SCIENTIFIC INFORMATION®
3501 MARKET ST., PHILADELPHIA, PA 19104

The Double Helix 40 Years Later: Joshua Lederberg's Personal Commentary About Its Impact on Basic Research

Number 39 September 27, 1993

An Insider's View of a Revolutionary Discovery

This year marks the 40th anniversary of one of the most important events in modern science—Watson and Crick's landmark discovery of the DNA double helix.[1,2] Few other basic research breakthroughs have had as profound an impact on the future course of scientific investigation. Few others have led to as many practical applications in medicine, pharmaceuticals, biotechnology, agriculture, and other industries. And few have achieved such widespread public recognition.

In a *JAMA* issue commemorating the discovery, Joshua Lederberg, president emeritus of Rockefeller University, contributed a personal commentary on the impact of the double helix on basic biomedical research,[3] which is reprinted below. He offers a rather unique view on the topic. As a Nobel laureate, he is an especially insightful and knowledgeable commentator on the interplay between genetics, molecular biology, and biomedicine.[4] And his deep interest in the history and sociology of science gives him a broad and well-informed perspective on the process of knowledge discovery.[5-7]

In the reprint that follows, Lederberg reviews the historic development of molecular genetics and the key advances that led to Watson and Crick's discovery. In addition, he surveys major DNA-based research trends since this discovery. This discussion draws an interesting distinction between DNA as an "informational duplex" and a "mechanical helix." As Lederberg points out, "The most novel features of DNA are associated with its duplicity, rather than its helicity."

The Double Helix: A Personal Note

In addition to its fundamental impact on basic research, Watson and Crick's discovery had an important benefit that is not often recognized. That is, it *humanized* science. This does not refer directly to the primordial 1953 *Nature* papers[1,2] but to Watson's 1968 book, *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*.[8] Twenty years later, Crick published his autobiography, *What Mad Pursuit: A Personal View of Scientific Discovery*.[9]

The Double Helix may not have been the first scientific autobiography, but it probably is the most prominent of the genre. One reason is that it frankly portrayed scientists as real people, not just idealized professionals. It showed they can be ambitious, competitive, and recognition-seeking and simultaneously objective, dispassionate, and disinterested.

Watson's book may have been one of the subconscious inspirations for the *Citation Classics®* feature in *Current Contents®*. I've often described these autobiographical commentaries as portraying the human side of science.[10] That is, authors are encouraged to informally recall the personal inspirations and frustrations of their

high impact work, rather than formally summarize the technical gist of the research. Some of these one-page commentaries certainly deserve to be expanded into articles or books. But the brevity of *Citation Classics*® commentaries is one of its main virtues. While some authors find it challenging to write these terse 500-word commentaries, most realize it is a special opportunity to call out the neglected aspects of their often unrecognized accomplishments.

In recent years, autobiographies by scientists in book form have become more common and popular. It would take considerable time and interest to read them all. But, minimally, short autobiographical accounts ought to become a standard practice for scientists. Indeed, I've often thought that a valuable innovation in scientific encyclopedism would be to ask all authors to combine short autobiographies with commentaries on their most-cited and/or even least-cited papers or books. Judging from the visible impact of the approximately 5,000 *Citation Classic* commentaries published to date, their content is already of considerable interest to sociologists of science.[11,12]

Watson, Crick, and Lederberg have not written *Citation Classic* commentaries about their Nobel Prize winning research. That might seem superfluous considering *The Double Helix, What Mad Pursuit,* or Lederberg's forthcoming autobiography. But as Josh has demonstrated below, it is difficult to believe that anyone has had the last word on these momentous discoveries.

*****

My thanks to Al Welljams-Dorof for his help in the preparation of this essay.

REFERENCES

1. **Watson J D & Crick F H C.** Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737-8, 1953.
2. ----------. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171:964-7, 1953.
3. **Lederberg J.** What the double helix (1953) has meant for basic biomedical science—a personal commentary. *JAMA—J. Am. Med. Assn.* 269(15):1981-5, 1993.
4. ----------. The interface of science and medicine. *Mt. Sinai J. Med.* 59(5):380-3, 1992. (Reprinted in: *Current Contents*® (21):4-8, 24 May 1993.)
5. ----------. Does scientific progress come from projects or people? Address to the National Association of State Universities and Land Grant Colleges. Washington, DC, 9 November 1987. (Reprinted in: *Essays of an information scientist: creativity, delayed recognition, and other essays.* Philadelphia: ISI Press, 1991. Vol. 12. p. 336-44.)
6. **Zuckerman H & Lederberg J.** Postmature scientific discovery? *Nature* 324:629-31, 1986. (Reprinted in: *Ibid.* p. 18-23.)
7. **Lederberg J.** Introduction: reflections on scientific biography. (Lederberg J, comp.) *The excitement and fascination of science: reflections by eminent scientists.* Palo Alto, CA: Annual Reviews, 1990. Vol. 3. Pt. 1. p. xvii-xxiv. (Reprinted in: *Ibid.* Vol. 13. p. 31-7.)
8. **Watson J D.** *The double helix: a personal account of the discovery of the structure of DNA.* New York: Atheneum, 1968. 226 p.
9. **Crick F H C.** *What mad pursuit: a personal view of scientific discovery.* New York: Basic, 1988. 182 p.
10. **Garfield E.** *Citation Classics*®—four years of the human side of science. *Current Contents* (22):5-16, 1 June 1981. (Reprinted in: *Op. cit.*, 1983. Vol. 5. p. 123-34.)
11. **Astin H S.** *Citation Classics*: women's and men's perceptions of their contributions to science. (Zuckerman H, Cole J R & Bruer J T, eds.) *The outer circle: women in the scientific community.* New York: Norton, 1991. p. 57-70. (Reprinted in two parts in: *Current Contents* (33):5-10, 16 August 1993, and *Current Contents* (35):6-11, 30 August 1993.)
12. **Campanario J M.** Consolation for the scientist: sometimes it is hard to publish papers that are later highly cited. *Soc. Stud. Sci.* 23:342-62, 1993.

Reprinted with permission from: **Lederberg J.** What the double helix (1953) has meant for basic biomedical science. A personal commentary. *JAMA—J. Am. Med. Assn.* 269(15):1981-5, 1993.

What the Double Helix (1953) Has Meant for Basic Biomedical Science

A Personal Commentary

Joshua Lederberg, PhD

THE ARTICLE published by Watson and Crick in 1953[1] was the landmark pointer to our contemporary model of DNA as a macromolecular structure. This lay on a well-worn path of biophysical analysis, reducing microscopic anatomy to the molecular level. It also helped inspire an enormous body of biochemical research that has defined DNA as *the* informational molecule, a discontinuity that has been labeled the Biological Revolution of the 20th Century. As a piece of structural analysis, the idea of the double helix includes the concepts (1) that DNA is a duplex structure, comprising two paired complementary strands, associated by secondary, noncovalent bonds; (2) that the strand pairs are coiled, forming a double helix; and (3) that these are antiparallel—the orientation of one strand being in the opposite polarity from the other.

The most novel features of DNA are associated with its duplicity, rather than its helicity. Linear polymers rarely form stiff straight rods; folding into coils is the norm. The genetic functions of DNA are inextricably associated with its duplex structure, and hardly at all with its helical shape; this is reflected in the preoccupation of DNA research with its role as an informational molecule. However, we shall see a recent concentration of interest in supercoiling. Inevitably, the biochemical interactions of DNA with other molecules, be they regulatory proteins or chemotherapeutic inhibitors, will often be intimately wound up with the precise three-dimensional conformation of the helix. This is also proxy for higher orders of coiling, interactions with histones and other DNA-binding proteins, and the organization of DNA into chromosomes.

Joshua Lederberg

DNA can be built in either an antiparallel or a parallel format, although the former adds a note of symmetry that may account for the prevalence of the antiparallel in nature. For parallel DNA a different enzyme would be needed to recognize and replicate from the 5′ compared to the 3′ end of the double helix. Recognizing this asymmetry, Watson and Crick[1] speculated that DNA was antiparallel prior to concrete observational evidence for this conformation.

Rarely has a structural determination been coupled so promptly with functional implications. Watson and Crick[1] immediately inferred that DNA duplexes were

From the Rockefeller University, New York, NY.
Dr. Lederberg is a consultant for Affymax Technologies, which uses DNA combinatorics.
Reprint request to The Rockefeller University, 1230 York Ave., New York, NY 10021-6399 (Dr. Lederberg).

formed automatically when each strand was replicated, and that this involved the assembly of nucleotides, one by one, complementary to the existing structure.[2] They overreached the mark by suggesting that this might be possible even without the intervention of specific anabolic enzymes, the discovery of which we owe to the prodigious labors of Arthur Kornberg and his school in the 1960s. But in imputing autocatalytic powers to the DNA double helix, Watson and Crick[1] might lay claim to having anticipated the enzymatic functions of RNA (if not DNA), an iconoclasm that earned the Nobel Prize in 1989 for Sidney Altman and Thomas Cech.

Despite the intellectual revolution initiated by Watson and Crick,[1] we might still ask the question, At what point was the welfare of any patient altered by specific knowledge of the double helix? This is a question I agonized over during the 1970s, and its first answer was perhaps the work of Y. W. Kan on the prenatal diagnosis of hemoglobin disorders, using DNA hybridization (1978). How rapidly we have moved in the interval is recounted by Caskey[3] in the companion article. Why did that take 25 years? One may simply point to the enormous edifice of contributory knowledge that now bridges the most reductionist aspects of DNA structure to pathological manifestations.

HISTORICAL BACKGROUND OF WATSON AND CRICK

The biological role of DNA was still enmeshed in controversy in 1953. Nucleic acids had been extracted from pus cells by Miescher in 1869, and from the beginning were associated with cell nuclei. These substances are now known to be macromolecules composed of a linear array of nucleotides joined by phosphodiester bonds. Cytologists writing in the early 1900s remarked on the association of nucleic acids with chromosomes and speculated that this basophilic material in chromatin might be the substance of genetic continuity. This brilliant anticipation was, however, submerged by a misleading observation, namely, the apparent loss of basophilia in the chromosomes of oocytes, leading E. B. Wilson (1925) to remark "That the continued presence of 'chromatin' [ie, basi-chromatin] is essential to the genetic continuity of the chromosome has, however, become an antiquated notion." We now know that these chromosomes become remarkably unraveled in keeping with their massive involvement in transcription, associated proteins then overshadowing the continuity of the DNA.

This skepticism was reinforced by the apparent monotony of DNA structure embedded in Phoebus Levene's first analyses of DNA. They contained only four constituent nucleotides—each comprising a phosphate group, a sugar, and one of the four bases: cytosine (C), thymine (T), adenine (A), or guanine (G). Within the limited analytical precision available in the 1920s, these appeared to be present in exact stoichiometric equivalence. Hence the provisional hypothesis of DNA as a tetranucleotide, although it was well recognized that its molecular weight and other key parameters had yet to be ascertained. Nor was there any biological system or array of sources to tell that one DNA preparation was in any way different from any other. Such a simple molecule seemed a poor candidate for the miraculous capabilities of the gene. On the other hand, proteins contained an abundant variety of constituent amino acids (eventually 20). More important, dozens, even hundreds of proteins were isolated with vastly different biological, physical, and chemical properties, including wide disparities in composition. The 1920s saw the most exciting developments in protein chemistry, even the crystallization of urease and of pepsin and the demonstration that enzymes were pure proteins (Sumner, 1926; Northrop, 1930). The cap seemed to be a similar characterization of the tobacco mosaic virus, claimed to be pure protein by Wendell Stanley in 1935. This was, however, soon to be corrected by Bawden and Pirie in 1937, who found phosphorus and carbohydrate in infectious concentrates of tobacco mosaic virus and inferred the presence of RNA. Stanley, nevertheless, received the Nobel Prize in chemistry in 1946,

together with Sumner and Northrop. By that time, Stanley acknowledged "that the nucleic acid could not be removed without causing loss of virus activity and there was general agreement that the virus was a nucleoprotein." Thus, this prize was a noble reinforcement of the primacy of proteins as the seat of biological specificity.

The breakthrough challenge to that dogma was thrust forth in 1944 by Oswald T. Avery, Colin MacLeod, and Maclyn McCarty. They had studied the diverse serological types of the pneumococcus and followed up Griffith's report (1928) that these could be altered or transformed by extracts of other strains. The gist of the 1944 study was that the transforming substance was DNA! This was contrary to expectations that the carbohydrate antigen or some associated protein would be the transforming substance. Avery, a member of the same Rockefeller Institute as Wendell Stanley, was intimately familiar and impressed with the difficulties of characterizing biopolymers. Though fully cognizant of the biological implications of the discovery, he was even more hesitant to dwell on them—but did include a remark that "The inducing substance has been likened to a gene...."

Their claims, of course, aroused intense critical controversy, largely around the obvious question whether their DNA preparations were still contaminated with traces of biologically active protein. Avogadro's number, $6x10^{23}$ per mole, would allow a residuum of 10^{7} protein molecules per microgram of a preparation that was 99.99% protein free, at the limit of analytical detectability. The sensitivity of the active materials to deoxyribonuclease might be ascribed to a protective rather than informational function of the DNA. Likewise, the insensitivity to proteases might be an attribute of a nucleoprotein complex.

My own role in the debate was a willingness, even desire, to believe—but a sense of responsibility that the issue was too important to be regarded as closed until there was no escape. It was not clear what feasible experiments (short of ab initio synthesis of DNA) could ultimately seal all these infinitesimal loopholes. One might go along with "DNA" as a working hypothesis, and some did. Most biologists blurred their judgments by talking about nucleoproteins—not necessarily informed by the distinction they were implying. Some might have meant something like "protein" or "nucleic acid" or a combination thereof, but please do not ask the role of the constituents. A rare few gambled on the DNA—as in some sense did Watson and Crick,[1] although they would have enjoyed working out its structure regardless of its biological implications. In the event, the final elucidation of DNA structure was a horse race. By Watson's own account, only a few weeks would have separated their priority from the looming insights of Maurice Wilkins and Rosalind Franklin (who had provided the critical experimental data) or of Linus Pauling.

The biological significance of the pneumococcus transformation was also problematical. It looked like a transfer of genetic information; but until 1951, the only markers tested were the serotype antigens. Could one extrapolate from those to genes in general, particularly given that the very idea of a bacterial genetics was in its infancy?

After the 1944 bombshell, more chemical attention was given to the tetranucleotide model, and signs of greater chemical complexity emerged. Of particular import were the deviations of the four bases from the simplistic 1:1:1:1 ratio, found by Erwin Chargaff. Furthermore, DNA from different sources exhibited different base composition. So perhaps DNA could be more complex, more diversified than previously thought—could be rehabilitated as a candidate for the gene. During the 1940s the Feulgen cytochemical test for DNA and analyses indicating constancy of DNA per genome in somatic cells and a halving in germ cells also added to DNA's respectability. But these findings did not necessarily prove more than a structural or scaffolding role for the DNA. The pneumococcus transformation remained the only biological assay for a genetic role for DNA—in contrast to the innumerable enzyme and immunological assays available for candidate proteins.

This impasse was alleviated by the broadening of phage research, sternly governed by Max Delbruck's genius, to embrace a wider range of chemical studies of phage infection. A critical one was the 1952 double-labeling experiment of Hershey and Chase. Most of the S-35 label (capsid protein) was excluded from infected cells; most of the P-32 (DNA) entered and was transmitted to the phage progeny. This experiment is often cited as the crowning blow on behalf of the "DNA-only" model. But Hershey himself did not go so far—well aware that "most" is not "all," he was still referring to "nucleoprotein" in 1953—and this at the same Cold Spring Harbor Symposium that sponsored a critical discussion of the paper by Watson and Crick.[1]

The article by Watson and Crick[1] did not, of course, bear directly on the loopholes in Avery's claims. It did add a further note of plausibility to a DNA-only concept of the gene. In the absence of any serious contradiction, this gradually hardened from working hypothesis to central dogma. The most serious challenge today is the prion hypothesis: that some "infectious" agents may be devoid of nucleic acid. This is still contentious at an experimental level: the hypothesis least in conflict with nucleic doctrine is that the infectious prion is a sort of epitaxial primer of aggregation of a host-determined protein. This still leaves obscure how and whether different prions could maintain and propagate their identity in a genetically defined host.

Long after many other lines of evidence converged to support an informational role of DNA—eg, the colinearity of DNA sequences with protein products (Yanofsky), genetically active DNA was eventually synthesized in the chemical laboratory (Khorana) and replicated enzymologically (Kornberg), fully vindicating Avery et al and those who gave their faith to these propositions.

THE FLOWERING OF MOLECULAR GENETICS

Since the rediscovery in 1900 of Mendel's 1865 work, genetics has had an extraordinary development, even without the benefit of tangible physical and chemical models of the genetic material. The biological phenomena of mutation and of sexual crossing (genetic recombination) opened the door to experiments in which existing organisms were the reagents. Genomes could be mixed by crossing, and new combinations of factors segregated into the offspring. Likewise, fruit flies could be subjected to radiation, and variant or mutant forms discovered. Genetic information is organized into linear chromosomes, and the processes of meiosis in gametogenesis: precise synapsis of homologues and crossing-over or segmental exchange of chromosome parts allowed powerful dissection of fine structure on a scale that rivals that of microchemical analysis. These methods continue to play an indispensable role in the denomination and mapping of mutant genes. By 1941, through the work of Beadle and Tatum, the groundwork of biochemical genetics had been laid—the role of genes in the prescription of protein products, and the use of mutations in the dissection of metabolic pathways. Indeed, many of these ideas had been anticipated by Archibald Garrod's studies of human biochemical defects at the very dawn of genetics.

Since 1953, we have had a new language for the description of genes: they are now segments of DNA that can be defined and manipulated as chemical entities. The linguistic transition has been conceptually smooth, though marked by occasional generational quarrels. Understandably, very few individuals can combine erudite knowledge of the life histories of a wide range of organisms in their natural habitats with focused and specialized knowledge of biochemical manipulations in the laboratory. Nor have many radical revisions of genetic doctrine issued from the molecular perspective. We have had to acknowledge that genes, as bits of DNA, are subject to a wider range of chemical and biological interactions than was previously thought—especially with other DNA. The icon of stability of genomes has been shaken by

the discovery of transposable elements, first noted in maize by McClintock in 1951; these remained inexplicable until they could be studied as DNA molecules. And concentrating on DNA now allows us to inject genes with viruses, needles, even "shotguns," into a range of cellular targets including the germ line, providing a technical revolution in the construction of new genotypes in all kinds of organisms—bacteria, plants, and mammals.

Meanwhile, other advances, notably the extension of recombination analysis to somatic cells in culture by cell fusion, have extended the technical power of genetic analysis in ways compatible with, but not dependent on, the double helix. It is paradoxical that the human chromosome number, $2n = 46$, was not correctly understood until 1956 (Tjio and Levan), and that for about 20 years thereafter this was at least as important in the development of human genetics as was the structure of DNA.

The adumbration of DNA-based research, molecular genetics, since 1953 would embrace a substantial fraction of world science. Many encyclopedic monographs struggle to record the details and promptly become obsolete. We can hardly do more herein than summarize the major headings, following an imprecise dichotomy distinguishing topological DNA—an informational duplex—from mechanical DNA—a three-dimensional geometric object.

DNA AS AN INFORMATIONAL DUPLEX

Denaturation and Hybridization

The most elementary aspect of the duplex is the separability of its strands, using temperature or chemical denaturants. A-T base pairs melt (separate from one another) at a lower temperature than G-C pairs, so melting curves can distinguish DNA of different base composition. Single strands once separated can also be reannealed, allowed to rejoin, the kinetics allowing the discovery that much DNA (in eukaryotes) has a repetitive or a redundant sequence. Radioactively labeled probes can be used to ferret out target homologous DNA with high precision.

Homology and Evolution; Polymorphism Within the Species

These and related methods can be used as quantitative indices of the genetic relatedness of diverse species, supplanting the subjectively evaluated morphological criteria used in systematics heretofore. Within the species, genetic polymorphism can now be described at the DNA level—one astonishing finding is that humans are typically heterozygous with a prevalence of two or three per 1000, ie, almost once in every gene. As most of these base substitutions have no perceptible phenotypic effect, random drift (rather than selectible or adaptive change) may predominate in evolutionary change (Kimura).

Mutagenesis and DNA Repair

The vulnerability of genes to mutational change in response to x-rays was known empirically since 1927 (Muller), and to chemicals since 1944 (Auerbach). Early hopes that chemical mutagenesis would be a direct path to the chemistry of the gene were not substantiated. Most chemical mutagens react with amino acids as well as DNA bases. The exceptions are nuclein base analogues, which may be misincorporated into DNA; but these were discovered much later. Above all, we now understand that the initial lesions in DNA would usually be lethal, and that eventual mutations are the result of intricate repair metabolism that occasionally misfires.

Transcription; Genetic Code

The "central dogma" of information flow has emerged, that DNA $\dashrightarrow$ (transcription) RNA $\rightarrow$ (translation) protein. The base sequence of DNA is transcribed faithfully into a messenger RNA copy. This in turn governs the assembly of a polypeptide sequence, each three-base frame of RNA encoding one particular amino acid. The polypeptide then folds (perhaps with the guidance of a chaperone) into a preordained protein three-dimensional shape, which can then function as an enzyme, antibody, hormone, structural unit, and so forth. This folding process is not yet fully computable. There may even be circumstances

where a given polypeptide might have alternative foldings—but this is not accepted dogma.

The detalis of messenger RNA synthesis have become much more intricate. Primary transcripts are usually processed, only some of the RNA tracts being spliced together to form the final message. The other "intervening sequences," or introns, may be the major part of the RNA—their functions remain obscure. As with repeated sequences, they may reflect "selfish DNA," whose presence in the genome has little to do with their adaptive value to the overall organism. In other examples, RNA may be edited in other ways before translation is completed.

Enzymology: Nucleases, Ligase, Replication; Reverse Transcriptase

For a legion of brilliant and tireless investigators, the DNA structural model has been the platform for isolating a host of enzymes involved in every aspect of DNA metabolism. Besides giving us that metabolic map, explaining how DNA is replicated, sliced, stitched, spliced, and repaired, these enzymes are the vital technical tools for further study of DNA and for the engineering of new constructs.

Some viruses, notoriously the retroviruses (including human immunodeficiency virus), exhibit reverse transcriptase, whereby RNA → DNA. This knowledge is indispensable to the virologist. It has also given some of the most valuable tools for studying RNA, eg, messenger, by allowing the production of DNA copies for input into other technology.

Tools for Engineering: DNA Splicing; PCR

These sempstering tools have founded the multibillion-dollar biotechnology industry. DNA tailored in vitro, with inserts from human or a variety of other sources, can be patched into convenient host garments (from bacteria to cows) for the easier exhibition of a variety of products—growth factors, enzymes, immunizing antigens, replacement therapeutics (like clotting factors)—in unlimited variety. Related technology is used to target specific host genes, to elucidate their functions in physiology and development.

The PCR (polymerase chain reaction) has been the instrument of the "democratization of molecular biology." With it a single DNA molecule in a messy mixture can be fished out and amplified ad libitum, most importantly at low cost and with simple instruments. High school students do experiments today that would have been doctoral dissertations 15 years ago. The applications range widely, from forensics and diagnosis of genetic disease to the hunt for new viruses and the revival of fossil DNA. At its heart, a synthetic DNA probe is a rational, linear, digital signature to locate any counterpart in the analysand. Its core of combinatorial specificity can be contrasted with that of antibodies, which is founded on three-dimensional shapes of the immunoglobulin and its targets.

Drug Discovery

DNA combinatorics has reached a new peak in a paradigm for drug discovery that mimics natural evolution.[4] Randomized DNA sequences are expressed on host cells (or phages), and these are then selectively screened for specificities of binding to specific reagents—usually receptors for which agonists or antagonists are sought. The cell expressing the desired epitope can then be grown out for larger scale production and testing. In one application, the mammalian antibody-forming mechanism can be emulated, and mutant immunoglobulin polypeptides selected for the desired specificity. RNA can fold into stereospecific objects; hence, randomized RNA molecules can be directly selected and replicated with reverse transcriptase.

Human Genome Project

With the availability of all of these tools, the image has firmed of establishing the complete DNA sequence of the human genome. As a scientific objective, this is uncontroversial. The controversy pertains to the primacy given to the staging of the effort. Should it be a once and for all technological production, mindless of the an-

cillary interest in some genes or DNA tracts compared with others? Does it need to be a centralized project, administered top-down with the trappings (and political appeal) of other Big Science? Or can it be left to the cumulative efforts of hundreds or thousands of laboratories, each digging more deeply at some features of the terrain, and intent on going much further than establishing a sequence of bases? In fact, we are seeing the emergence of constructive compromise among these visions; and at the same time the technologies of mapping and sequencing are advancing to where the costs of a unified project need no longer prejudice more individualized efforts.

In any case, sequence information is but the beginning of more intensive inquiry into the polymorphisms, regulatory factors, and gene functions associated with any DNA segment.

DNA AS A HELIX

Higher Orders of Organization

The visible chromosome is a packaging of DNA, histones, and accessory proteins three or four orders of coiling beyond the double helix. Cytological observation leaves no doubt that the morphological expression of the chromosome reflects functional allocation of different genes; but we are at the mere beginning of understanding.

Gene Regulation and Morphogenesis

The basic outlines of the central dogma now consensually agreed, the core challenge of molecular biology has been the path from the gene to the organism. Given that, to some approximation, each somatic cell has the identical genotype, (1) how is gene expression differentially modulated, and (2) how is this transmitted in cell lineages?

A multitude of DNA-binding proteins have been found that do modulate gene expression: transcriptional regulators. As a three-dimensional interaction, protein binding is fully sensitive to three-dimensional shape and the major and minor grooves of the double helix, as well as the base sequences contained therein. In addition, if not in consequence of bound proteins, some tracts of DNA are methylated shortly after DNA replication, in ways correlated with gene activation.

How these properties are locally transmitted remains a matter of speculation, but may well be bound up with local methylation.

DNA Supercoiling; Topoisomerases; Other Conformations

The standard double helix exhibits a pitch of about 10 base pairs per complete turn. If nothing else, the processes of replication and transcription would entail the unraveling and rewinding of the helices: this is the task of enzymes generically called topoisomerases. These can transiently cut single strands to permit the relief of torsional stress, then rejoin them. In its natural habitat, DNA is often found in states of positive or negative supercoiling, often correlated with maintained gene expression. In addition, many cytotoxic and cancer chemotherapeutic agents seem to be topoisomerase inhibitors, and most owe some of their specificity to the momentary DNA-supercoil status of a given cell. It is particularly intriguing that environmental signals can modulate that status, often by regulating the production of the various topoisomerases.

At least in vitro, DNA can undergo a spontaneous transition to a totally different, kinked and left-handed conformation called Z-DNA. Tracts rich in G-C pairs are especially prone to this shift. The importance of Z-DNA in vivo is hotly contested.

DNA conformations plainly confer different chemical reactivity on the bases, a principle exploited by the footprinting methods used to study conformation. This must have some implications for localized chemical mutagenesis—a matter not yet systematically studied.

TRIUMPH OF MECHANISM

The dominion of the DNA paradigm has been the triumph of mechanistic interpretation in 20th-century biology. It is sometimes remarked that human personality is nothing but the individual's 3 billion base

pairs—an assertion that fascinates some, terrifies others, and has much to do with the debate about the Human Genome Project. If we could believe that existing genotypes had achieved more than a tiny fraction of the human potential—in culture, in intellect, in compassion, in a sane ordering of affairs—we could elevate the genome to that pedestal of nemesis. On the other hand, we do know that many, probably most, individuals labor under some potentially remediable burden of hereditary origin. As much to understand the better nurturing of human development, a euphenics, as to intervene in genetic constitution, eugenics, it does behoove us to learn all we can about genetic polymorphisms and their impact on human health and capability. It is particularly important to distinguish interventions in germ cells from those in the somatic cells, and to communicate that it is only the latter that are intended to be the targets of the new gene therapies.

SELECTED READINGS

It would be a precious exercise to provide specific documentation for every detail of this commentary; it would be both arduous and redundant—many single points would deserve a library. The up-to-date detail can be found in standard texts of molecular biology (a few are listed) and in the volumes of *Annual Review of Biochemistry*. The leading historical monographs on DNA are also listed.

Molecular Biology

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. *Molecular Biology of the Cell.* New York, NY: Garland Publishing; 1989.

Berg P, Singer M. *Dealing With Genes: The Language of Heredity.* Mill Valley, Calif: University Science Books; 1992.

Darnell JE, Lodish HF, Baltimore D. *Molecular Cell Biology.* New York, NY: Scientific American Books; 1990.

Kornberg A, Baker TA. *DNA Replication.* 2nd ed. New York, NY: WH Freeman & Co; 1992.

Stryer L. *Biochemistry.* New York, NY: WH Freeman & Co; 1988.

Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. *Molecular Biology of the Gene.* Menlo Park, Calif: Benjamin/Cummings; 1987.

Wells RD, Harvey SC, eds. *Unusual DNA Structures.* New York, NY: Springer-Verlag NY Inc; 1988.

History of DNA

Crick FHC. *What Mad Pursuit: A Personal View of Scientific Discovery.* New York, NY: Basic Books Inc Publishers; 1988.

Dubos RJ. *The Professor, the Institute, and DNA.* New York, NY: Rockefeller University Press; 1976.

Fruton JS. *Molecules and Life.* New York, NY: Wiley-Interscience; 1972.

Judson HF. *The Eighth Day of Creation.* New York, NY: Simon & Schuster, 1979.

Lederberg J. Genetic recombination in bacteria: a discovery account. *Annu Rev Genet.* 1987;21:23-46.

McCarty M. *The Transforming Principle.* New York, NY: WW Norton & Co; 1985.

Olby RC. *The Path to the Double Helix.* Seattle: University of Washington Press; 1974.

Portugal FH, Cohen JS. *A Century of DNA.* Cambridge, Mass: MIT Press; 1977.

Sayre A. *Rosalind Franklin and DNA.* New York, NY: WW Norton & Co; 1975.

Watson JD. *The Double Helix.* New York, NY: Atheneum Publishers; 1968.

Watson JD, Tooze J. *The DNA Story.* San Francisco, Calif: WH Freeman & Co; 1981.

References

1. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature.* 1953;171:737-738.

2. Watson JD, Crick FHC. Genetical implications of the structure of deoxyribonucleic acid. *Nature.* 1953; 171:964-967.

3. Caskey CT. Molecular medicine: a spin-off from the helix. *JAMA.* 1993;269:1986-1992.

4. Pei D, Ulrich HD, Schultz PG. A combinatorial approach toward DNA recognition. *Science.* 1991; 253:1408-1411.