Speyer, J. F., Lengyel, P., Basilio, C., and Ochoa, S. (1962). *Proc. Natl. Acad. Sci. U.S.* **48**, 63, 441.

Stevens, A. (1961). *J. Biol. Chem.* **236**, PC 43.

Volkin, E., Astrachan, L., and Countryman, J. L. (1958). *Virology* **6**, 545.

Warner, R. C. (1957). *J. Biol. Chem.* **229**, 711.

Watson, J. D., and Crick, F. H. C. (1953). *Nature* **171**, 964.

Weiss, S. B., and Nakamoto, T. (1961). *Proc. Natl. Acad. Sci. U.S.* **47**, 1400.

Wood, W., and Berg, P. (1962). *Proc. Natl. Acad. Sci. U.S.* **48**, 94.

Zamecnik, P. C. (1960). *Harvey Lectures, Ser.* **54**, 256.

A "Book Model" of Genetic Information-Transfer in Cells and Tissues[1]

JOHN R. PLATT[2]

Department of Physics, University of Chicago, Chicago, Illinois

Abstract

The expression of genetic information in cells and whole organisms is like the reading out of a complex instruction manual, but the analogy extends to more details than is generally realized. The information is linearly arranged in "words" that are "read out" sequentially in time. There is one copying mechanism (DNA polymerase) for reprinting the whole book, and another (RNA polymerase) for selective read-out into cell chemistry. The read-out is by "paragraphs" (genes) and by "pages" (operons) that can either be "closed" (repressed) or "opened" (induced), according to contingent "instructions" (repressor-corepressor complexes) from "references" (regulator genes) on earlier pages or in the "books" of adjacent tissues.

One of the most important implications is that the "pages" are "referred to" by "page numbers" or "page headings" (inducers for operator-genes) that must be widespread low-information codes very different in kind, and different chemically, from the rest of the species-specific high-information "text" (structural-genes) on the "page." Between "readings," the "books" can be "locked away" in compact "storage" forms (phage heads, chromosomes). This analysis points up several questions, such as the "reading" of the two strands in a double-helix, the many roles of protein-nucleic-acid "translator" molecules, and the diverse mechanical configurations of the chains as they are "opened" and "closed" and "stored."

I. Introduction

The presently known facts of genetic information-transfer have been reviewed by Jacob and Monod (1961) and summarized by Rich (1962a). Let us recapitulate the main results, as described in the current genetic-biochemical language.

"DNA makes RNA makes protein." That is to say, the long hereditary deoxyribonucleic acid (DNA) chains [or sometimes ribonucleic acid (RNA) chains], containing specific sequences of bases, can be copied in short sections by Messenger-RNA chains; which are then copied (or, better, "translated") into the amino-acid sequences of enzymatic proteins.

What determines whether a particular enzyme is manufactured or not?

Jacob and Monod show that a "regulator-gene," perhaps distant from the enzyme locus, makes a "repressor-substance." This substance can complex with certain cellular constituents called "co-repressors" or "inducers." It then interacts, or fails to interact, with a particular "operator-gene" on the DNA chain so as to "repress" or "derepress" (that is, "induce") the manufacture of Messenger-RNA from neighboring "structural-genes," among which is the gene determining the structure of the enzyme in question.

Several enzymes, frequently those involved in successive steps of the synthesis of the same cellular product (histidine, for example, or arginine, or orthophosphate), may be under the control of the same repressor-gene. In bacteria, they are often clustered in sequence on the genetic map and are under the control of a single operator-gene as well. An operator-gene and the enzyme structural-genes

that it controls, together make up a single information unit called an "operon."

In this picture, the co-repressor that acts to suppress the formation of an enzyme is frequently the excess product (histidine, for example) of the reaction-sequence that the enzyme takes part in. On the other hand, an inducer that acts to get an enzyme formed is frequently the substrate (glucose, for example) of a degradative process that the enzyme mediates. In short, the system has "feedback control" to keep enzymes from being produced unnecessarily, except when mutations damage this regulatory system. Since most enzymes are needed only on specific occasions or for specific contingencies, their operons stay most of the time in the "repressed" state.

The present remarks are a series of reflections on these curious and interesting relationships. They were inspired by the realization that there is a simple and familiar analogy to any chain of information that is expressed, section by section, in this all-or-none fashion. It is the everyday example of the information-sequence in a book—or more precisely, perhaps, a very complex "instruction manual"—that can only be open to one page, or a few pages, at a time.

For many years, of course, writers and speakers on genetics have used the linguistic and bookish metaphors of "codes" and "commas" and "words" that can be "read out." But the point here is that this casual analogy can apparently be extended now to many additional details of genetic storage and read-out. It therefore becomes interesting to discuss more carefully the question of just how exact the analogy is, and at what points it begins to break down. If it is accurate in a sufficient number of details, a serious comparison between the genetic chain and an instruction book could become a useful device for teaching and for explanation in genetics and embryology, it could give us a more general and coherent view of the genetic process, and it could point up research questions concerning structural and biochemical analogies that we might otherwise neglect.

We can see these possibilities better by examining the various parallels one by one. We will begin with the most familiar ones.

II. One-Dimensional Sequence of Information

The information in a book (except for the pictures) is ordered in a one-dimensional array; so is the genetic information. From the

beginning, Morgan's hypothetical "genetic map" was assumed to be linear in order to represent the experimentally transitive relations among the recombination probabilities of the genes. Benzer has now proved conclusively the linearity of the microgenetic map has now proved conclusively the linearity of the microgenetic map within a gene (the r_{II} section of T2 phage) all the way down to the level of the point-mutations (single base changes) (Benzer, 1959). Electron-microscope pictures of DNA chains deposited on a backing also show that they may extend unbranched for thousands or tens of thousands of angstroms (Beer, 1961).

It is true that the genetic map in T2 and T4 phage appears to be circular, but this may be a purely formal effect of some kind of continuous replication rather than the result of physical looping of the DNA (Stahl, in press). A pulling-out of a helix or double-helix chain into side branches during some part of the life cycle is also conceivable (Platt, 1955), but it would not change the linearity of the genetic information-sequence. Consequently, although our present picture of the genetic chain is more like the continuous line of writing on a scroll than the divided lines of a Western book, the one-dimensional information-relation in either the book or the scroll is evidently an accurate representation.

III. Sequential Read-Out

It has now been shown that the amino-acids are not added to the Messenger-RNA template at random positions before uniting to form a protein, but are added sequentially from one end of the sequence to the other (Bishop et al., 1960; Dintzis, 1961). It has also been shown that the addition or deletion of a single base in the DNA chain shifts the amino-acid "read-out" forward or backward by one base for dozens or hundreds of bases "to the right" on the genetic map, which indicates sequential read-out in long sections within a gene at the DNA level also (Crick et al., 1961) (though of course, some sections might be read "to the right" and others "to the left").

As many authors have said, this is like reading out a special kind of writing, in which the "letters" (the 4 different bases in a DNA or RNA chain) are grouped, probably 3 at a time, into "words" (the 20 or so different amino-acids); but where the words are run together with no spaces or commas between them, so that a displacement by one letter backward or forward makes all the words "afterward" be read wrong.

To force the book analogy physically close to such a scheme, we would then have to do without separate lines or spaces or punctuation on a page. The letters would have to be printed in one continuous string from beginning to end, perhaps winding back and forth across the page (if this is necessary for compactness) in the old "boustrophedon" style that is still found sometimes in childrens' reading-puzzles. But a normal book gives us this same directional and sequential continuity so far as the sense is concerned.

IV. One Line to Read (Not Two)

In the Watson-Crick double-helix picture of DNA, each of the two complementary strands carries an information sequence that could be read out into one of two complementary Messenger-RNA sequences. Are both of these Messenger-RNA's actually formed? If so, does each make a different protein, of the same length but with different and useful functions? (Rich, 1961, 1962b). The latter, at least, has always seemed highly unlikely to those who have considered the question, and there is now direct chemical evidence against it.

For example, Speyer et al. (1962) and Matthaei and associates (1962) have reported the synthetic RNA base-triplet "words" that "make" the different amino acids and have found that the "meaningful" triplets are all non-complementary to each other with, at most, two or three exceptions. The meaningful triplets all contain a U, or are U-rich; the complementary ones would necessarily be U-poor.

These results point to a series of conclusions, as suggested by Rich (1960, 1961). One is that most or all of the complementary RNA base-triplets (the U-poor ones) may not make amino acids. If so, in any section of a DNA double-helix, only one strand would need to be read, say the meaningful strand (U-poor in DNA) that gives a U-rich Messenger-RNA chain. (Note that on the one-strand-read-out assumption, it was never necessary to consider more than half the base-triplets—that is, 32 out of the possible 64—for coding the 20 amino acids.)

The DNA strand that is not read (or is read into non-protein-forming Messenger) should still not be regarded as "nonsense," but as something that could be called "compsense." For it still contains the information sequence, although in a complementary form that would require an additional replication (in a given generation) be-

fore it could be turned into protein. On the other hand, its presence eliminates one replication step between generations before daughter protein can be made; and it undoubtedly helps protect the other strand from breakage or chemical change. (On the book model, like the thin glazed sheets sometimes put in to protect expensive "illu-minated" pages.) The single-strand phage, ϕX-174, and the RNA viruses have shown us for some time, of course, that the second strand is not informationally necessary to genetics.

There is a difficulty in assuming that one DNA strand is a "sense" strand and the other "compsense," if "sense" means only the U-poor triplets. For this would imply (Rich, 1961) that the strands should then show easily detectable differences under electro-phoresis or centrifugation, contrary to what has been reported. But these results could be reconciled if each of the strands contain alter-nate sections of "sense" and "compsense" which tend to balance the purine-pyrimidine ratios. A similar suggestion was made earlier, on other grounds (Platt, 1955). In evolution, a "sense" section might become joined to a "compsense" section through failure of the second strand to separate from the end of the first strand at some replication (see also Rich, 1962a).

Messenger-RNA chains might terminate naturally at the end of a "sense" section, which would thus provide a natural end to a gene or a protein. If each strand "points the same way" (in each base-ribose-phosphate unit) for its entire length, but opposite to the other strand, as the Watson-Crick model of DNA suggests, then the "sense" sections on one strand would have to be read out "to the right," while the intervening sections where the "sense" is on the other strand would be read out "to the left." But both could be read out by the same read-out enzyme or RNA polymerase.

In any case, whether the read-out in every section is always in the same direction or not, we see that the genetic chains are close to the book model in that they form only a single linear sequence of sense, and are to be read out from one chain only, and not in duplicate, at each point. [Is the rate of read-out in different organ-isms also roughly constant, as it is for different books? Possibly. At least the lengths of the lifetimes required for a phage, a bacte-rium, and a man to develop all of their genetic information bear a very rough order-of-magnitude relation to the total lengths of their genetic DNA chains (U. Liddel, personal communication, 1961).]

V. Copying without Reading, and Vice Versa

We can copy a book—photographically, for example—without reading it. But monks and stenographers have also copied many a manuscript, letter for letter, without "realizing the sense" of it. In genetic information, too, we see that the "copying" process does not involve "reading." In copying, the enzyme DNA polymerase can go along replicating the two strands of a DNA double-helix from one end to the other, like a "zipper" (Rich, 1962a). No Mes-senger-RNA or protein is made. The whole chain is accessible to this enzyme—although it may be in this condition only during part of the life cycle, perhaps after its accessibility to "reading" has been blocked or finished. (Copying a book, too, interferes with reading it.)

Conversely, we can read a book, or sections of it, without making a permanent copy. "Reading" DNA chains "to get the sense out" requires a different enzyme, RNA polymerase, which goes along the chain making a complementary Messenger-RNA chain. This en-zyme, like us, does not read everything, but only those accessible or "derepressed" sections that have been "opened" for it. And it does not make a permanent copy, since the Messenger-RNA is short-lived and seems to be destroyed after making its protein. (The protein, of course, is *not* a "copy" of the instructions in the present sense. It cannot make further copies of itself, for example. It is instead the "working equipment" that the instruction manual shows how to assemble and set running.)

VI. Pages Open or Closed

The genetic book is not a "scroll," which can be opened for reading to sections of any length, with any beginning or end points. The evidence for "induction" and "repression" of clearly marked-off sections requires us to think of it at this point as a Western "book" with separate "pages" or, better, "double-pages" which are definitely opened or definitely closed at any given time (Jacob and Monod, 1961).

A "double-page" thus corresponds to "a *unit* in the transfer of information"—an "operon"—which is "opened" or "shut" by the action of an "inducer" or a "repressor" on a given "operator-gene." It may contain a single "operator-and-structural-gene," or it may contain an operator-gene together with two or more distinguishable

structural genes, governing two or more separate proteins (Jacob and Monod, 1961).

We might think of each of these structural genes as a "paragraph." (The paragraph even has "sentence" or "clause" subsections, the genetic "cistrons," which make protein fragments whose sections are complete enough to produce successful catalysis by mutual "complementation.") The page labeled "Histidine synthesis" in the *Salmonella* book, for example, seems to contain some 8 paragraphs of this kind, with the structural instructions for producing 7 different proteins that catalyze consecutive steps in the histidine synthesis (Jacob and Monod, 1961). (The eighth paragraph is the "heading" or "operator-gene.") The page can be closed, and all of these structural read-outs stopped, by excess histidine. It does not become partially closed; the process usually seems to have an "all-or-none" character. A page is "open" or "shut"; which is why, as we know, a vertebrate cell can do kidney chemistry or cartilage chemistry, but not something graded in between.

The lengths of pages and books. The known enzyme protein units seem to contain 100 to 500 amino-acids, or the same number of base-triplet "words" on their DNA chains. A "page" with 1 to 10 paragraphs may then contain of the order of 300 to 3000 words, say 2000 as a round number. This is comparable to the page lengths in large books or encyclopedias.

A bacteriophage or virus with a DNA chain, say, 200,000 bases long, then has about 60,000 words or roughly 30 pages in its genetic instruction book. The book for a bacterium is 10 to 100 times larger. And for a man, the genetic information in the 46 chromosomes of each somatic cell is not so much a book as a very large encyclopedia with 46 volumes, about 6×10^9 base-pairs, 2×10^9 words, and a million pages; or an average of about 20,000 pages per volume. The reason for dividing large books into many separate volumes, in the biological as in the literary case, might be that it makes all this information mechanically easier and faster to reprint and handle. Also it may be mechanically easier, if many pages need to be open at the same time, to have them open in separate volumes.

VII. Only a Few Pages Are Open

A cell is particularly like a book not only because the pages can be opened or closed as needed, but because only a few pages need to be open at a given time. For one thing, different enzymes are

wanted for different environmental conditions, for example in a cell that can subsist on many different sugars. As genetic information, they may represent evolutionary adaptations that have been useful in the past, but only a few of them may be needed to survive on a particular substrate. Many enzymes also may be needed to make certain products only at one stage of the life cycle and not at another.

Jacob and Monod correspondingly emphasize the need for repressive regulation of most enzyme-making genes. They point out that "constitutive mutants," in which the control of a particular enzyme has broken down, may turn out 6% or more of their protein in the form of this enzyme (not the enzyme-product, but the enzyme!). Obviously "cells could not survive the breakdown of more than two or three of the control systems." That is, the cell cannot "hold open and read out continuously" more than two or three of its hundreds of pages.

The limitation on the fraction of its information that a cell can express at a given time is even more dramatic in a multicellular animal. Clearly the ovum, before its first division, cannot express the full informational content of the animal, and it is generally recognized that most of its genes are not functioning. Likewise in the adult animal, no single cell is expressing the information represented by all the other types of tissues. The information is "there" (at least in the somatic cells of higher animals) but it is repressed. This is proved by the classical embryological experiments on the "induction" of cells to form some of the other tissues that they had not been "destined" for (skin into eye, muscle into cartilage).

Recently, in fact, it has been recognized that this selection of only a part of the genetic information to be active chemically at a given time can be seen under the microscope. The "Balbiani puffs" and swellings at particular spots on the stretched-out chromosome chains (in insects) are now known to be sites of synthetic activity (Beerman, 1959). And different sites along the chromosome are found to be swollen and active in different tissues, or in the same tissue at different stages of development.

A human somatic cell contains a few hundred times the DNA of a large bacterial cell. If it can do at one time about the same number of different chemical syntheses as the bacterial cell can do, then it must be using a hundredfold or thousandfold smaller fraction of its total genetic information. (It may be no accident that

the total number of differentiated cells that have so far been distinguished in man is of the order of a hundred.) We begin to realize that it would probably be energetically and chemically impossible for a tissue cell to use all the different kinds of information that the whole animal uses, unless the tissue cell itself became as large and as structurally differentiated as the whole animal.

The books in each cell may be open to only a relatively few pages at any one time.

VIII. Page Headings for Reference

We now come to a feature of the book analogy that has not received much genetic or biochemical emphasis. It is that every page of a book, except a very short book, always contains *two* different kinds of information:

(a) A long sequence of words, highly specific to the book, and so containing a large amount of information—that is, the text.

(b) A short sequence, common to many books, containing only a little information—that is, the page number or descriptive reference heading.

The specific text and its paragraphs we have seen in the structural genes. "Page number" or "reference heading" is probably a good metaphor for the operator gene. The way in which a regulator-gene at one place "refers" with its repressor-substance (if the right inducer is present) to an operator-gene to open an operon page, seems closely analogous to the steps and elements involved in a "conditional-transfer" instruction in referring from one page to another of a complex instruction book.

An example may make this clearer. As we are reading along on a certain page of our instruction book, we might come to the "conditional-transfer instruction" (that is, the regulator gene) that gives, in effect, this sequence of directions:

"See if histidine is present." (That is, let the RNA-polymerase or some other repressor-read-out enzyme make the repressor-substance-for-histidine from this regulator-gene information chain.)

"Then if histidine is not present" (that is, if the repressor substance does not now form a complex with any free histidine),

"turn to 'page 137,' the 'histidine-synthesis' page" (that is, let the uncomplexed repressor find the operator-gene that it fits—in its uncomplexed form—in some complementary way),

"and start making histidine from the instructions there" (that

is, let the RNA-polymerase get past this "derepressed" or "open" operator-gene and start peeling off the 7 kinds of Messenger-RNA for the histidine-synthesis enzymes from the 7 paragraphs of instructions).[3]

In this analogy, perhaps it makes little difference whether we think of the operator-gene as a page number or as a reference heading. (In computer-programming terminology, it is an "address.") The "number" idea might be more appropriate whenever the reading is "sequential" from page to page down the chain; for example, perhaps, in the first sections read out after mitosis or fertilization. (Though a chain that is read exactly sequentially, without any forward or backward referencing at all, would not even need page numbers.)

On the other hand, in most cell chemistry, and certainly in multicellular organisms, the pages will not be opened sequentially but according to function, and the operator-gene must then play a role close to that of a "short reference heading." Short, of course, because the operator-gene-code, like the reference heading, must be written in a language common to many organisms—the language of the common inducers and co-repressors—or else it must be "translated" by the repressor-substance into such a language. This means that the "recognition-section" of the repressor-substance—and the section of the operator-gene, also, that recognizes this repressor—only needs to be a short, "low-information," "heading" that responds to, and corresponds to, "histidine" or "glucose" or some other one of a relatively small number of widespread substances.

Thus a book—or a bacterium— with 900 pages can specify which page is open by using a number (an "address") with 3 digits or 10 bits, even though each page contains 1000 to 5000 bits of information. And each of these 3-digit numbers, or each of the 2-word page headings written on the same line with them in the anatomy book, is common to many books.

---

[3] The reader may be puzzled or amused by my choice of verb forms in writing down these "directions." They are chosen to represent the fact—the deviation from the "human-reading-a-book-model"—that there is no extra "little man" in the cell to "do" the reading or interpreting of the DNA instructions. The cell "reads itself," where "read" has become an intransitive verb, and means that the reading is essentially synonymous with the action of "turning pages" and "carrying out the instructions read." We have no customary verb forms in English for this "automatic" reading-which-is-action, in which the book does not require an external reader, and is not read, but reads.

This low-information feature must continue to characterize the more complex biological "inducers" by which one tissue affects another, and the repressors that recognize them, even though we do not know so much in detail about how each one acts. The hormones, for example, are one class of inducers. They are small and diffusible molecules common to many species, producing such similar tissue-development reactions that they must have low-information and low specificity compared to the structural proteins. The same is true of the more obscure tissue-inducer-chemicals. Structures such as eye-elements or feathers, of types specific to a given host-species, can be induced from the skin of the host in the wrong places, by embryo-transplants of inducer-tissues from an entirely different species or even class of animal. The structural genes induced are specific to the host; the inducer-chemical is specific only to the type of inducer-tissue. The inducer-chemicals, whatever they are, must therefore be common to many species or classes.

A similar organ-specific, low-information interpretation of a more subtle interaction seems to be demanded by the experiments of Moscona (1961a,b) and Weiss on reaggregation of separated ("trypsinized") tissue cells. They showed that chick kidney cells and mouse kidney cells "recognize" each other "as kidney" (at least at first, before any immune-reactions set in). The cells, meeting at random, form bonds with each other and then form organized tissue, reshaping and rearranging themselves into a hybrid pseudo-kidney, with kidney-like tubules and tubule-walls and cilia and other details. By contrast, when separated kidney cells are mixed with cartilage cells, either from the same species or from different species or classes, they "reject" each other and form separate kidney and cartilage tissues. This suggests assuming that some diffusible and widespread, but organ-specific, "kidney-recognition" molecules can diffuse into cells that have once been opened to the kidney "pages" and can open again the synthetic page whose "heading" reads: "Kidney, tissue formation with neighbors." But the other, "cartilage-recognition" molecules, diffusing in, are unable to open any synthesis at that heading. (Possibly they are blocked even from diffusing in; but this would simply involve the sub-case of the formation of "recognition-permeases," which is probably a complication that we do not need to go into here, since it seems likely that its genetic aspects do not involve any further new principle.)

Chemistry of "page heading" and repressor molecules. Evidently,

the "repressor-substance" must "recognize," and be capable of specific links with, *two* kinds of molecular groups: the operator-gene, and the co-repressor or inducer chemical. This suggests that it must itself have two functional groups, one of them probably protein, since the repressor is known to be stereo-specific for the inducer or co-repressor. It is true that the repressor-substance, at least during formation, does not test for protein, but it is hard to reconcile the stereo-evidence with any other alternative; and Jacob and Monod conclude that perhaps "the repressor itself synthesizes the 'induction protein' and remains thereafter associated with it."

By implication, the other, first, functional group in the repressor-substance could then be a nucleic-acid chain, like a Messenger-RNA that is not destroyed. It might be complementary in base-sequence to the regulator-gene DNA from which it is made. It might also be a template for the protein it becomes attached to. Or it might be complementary in base-sequence to the operator-gene DNA it can become attached to; or conceivably it might be all of these, although that is perhaps too much to expect. (On the other hand, an identity of both the regulator-gene and the operator-gene base-sequences would be like the identity of the two page-numbers on the line where the page is referred to and on the heading of the page itself.)

But whatever the details, this indication that the repressor-substance or "page-heading translator" is a compound of protein-plus-nucleic-acid, suggests immediately some important biochemical generalizations into which a number of other pieces of evidence can also be fitted.

For it is clear, on thinking about it, that if we are to have two different kinds of codes related to each other in cellular biochemistry—the base-sequence codes of the nucleic-acid chains, and the amino-acid-sequence codes of the protein chains—we must have, at every point of interaction between them, "translator molecules" containing *both* codes and able to "speak" both languages. The repressor-substances are molecules of this sort that translate from the inducer-code (whether all inducers are proteins or not) to the operator-gene DNA language. (If we suppose that their interaction is not with the operator-gene DNA but with some unknown "operator-gene protein," it only pushes the translation problem back onto this protein. In fact we see that when the repressor-substance is attached to the operator-gene, it *is* operator-gene protein.)

The general translation requirement suggests that we look for

other sets of protein-nucleic-acid translation-molecules. Two genetic-translation sets come to mind immediately. One is the set of Transfer-RNA molecules ("soluble-RNA"), each of which consists of an RNA-chain attached to its specific amino-acid. These are the molecules that transfer the base-sequence information from a Messenger-RNA template into a protein chain (Rich, 1962a). Another set is the set of postulated transfer-enzymes that "read" some base-sequence on these Transfer-RNA molecules in order to attach the proper amino-acids to them.

[It is instructive to spend a moment inquiring into the possible problem of "infinite regress" here. Will another second-order set of transfer-enzymes (or "language-teachers"?) need to be postulated to "read" the bases on the first set and attach the right amino-acid groups to them? And a third-order set to read these, and so on? To state the problem is to see it is absurd—and that it leads us to a chemical conclusion! For clearly at some stage, there must be either a chain of bases that does not need a further transfer-enzyme but that complexes spontaneously with a particular amino-acid; or a chain of amino-acids that does the same with a particular base. Where does this stage occur? The translation from RNA base-triplets to amino-acids already appears to be general for several organisms (Nirenberg and Matthaei, 1961; Speyer et al., 1962). Therefore the Transfer RNA's themselves—or at most, their postulated transfer-enzymes—may already be universal. This can only mean that their base-and-amino-acid relationships are founded, like base-complementarity in the nucleic-acids, on steric and chemical reasons for "fitting," rather than on the accidents of biological survival. This in turn suggests a novel conjecture that seems strange but that might be worth exploring. A Transfer-RNA molecule contains some 80 bases but needs only 3 of them to "fit" the Messenger base-triplet; is it possible that this extra length is needed to make it a kind of "nucleic-acid-antibody" having a specific wrap-around relationship with its own amino-acid and with the other groups it must be fitted to? This would eliminate the need for a further transfer-enzyme.]

There may be another important set of protein-nucleic-acid "translator" molecules, namely those involved in embryological induction. There seems to be general agreement now that the "inducers" that diffuse from one cell to another are very likely RNA-protein compounds. Lash (1962) and Hommes et al. (1962) analyzed

chick-embryo spinal-cord inducer for cartilage-information in the somites. They found that it contains monophosphates of the bases guanine and cytosine, plus a hexose, a hexosamine, and some 15 amino-acids.

An embryological "inducer" acts in many ways like an "order" to a cell to "open" to a particular "heading" even if it is doing something else. This feature, together with the presence of the nucleic acid, suggests that it may not be like the small substrate-inducers complexing with a "repressor," but that instead it may be itself a repressor-competitor. If this were so, the need for endogenous "translation" of the exogenous inducer-compound would vanish. Certainly such a "translation-service" could often be eliminated between the cells of a species. And clearly, in that case, the interspecific generality found for tissue RNA-inducers must indicate an interspecific generality of the "operator-gene codes."

Whatever the nature and distribution of these various "translator" molecules, the translation requirement seems to demand that there be appreciable chemical and physical differences between regulator-genes, operator-genes, and structural-genes. To generate nucleic-acid-protein repressor-substances, for instance, demands some special kind of read-out of the regulator genes. Can this be done just by a special kind of base-sequence, or is a different kind of read-out-polymerase needed? This read-out might have to be done early in the life of a cell so that these "repressor-headings" could be attached to the operator-genes before any structural read-out is ready to begin. Does this mean that regulator-genes and operator-genes are specially accessible? If many regulator-genes have to be read out in this way "before the cell needs them," these repressor-syntheses might represent a major fraction of the early synthetic activity of the chains. The operator genes of course must have a different kind of chemical specialty; for they must evidently say "open" or "closed" to the RNA-polymerase that makes Messenger-RNA, but they need not themselves be copied into Messenger-RNA.

In the light of these considerations, it would not be surprising if the regulator and operator genes are marked off from the other information in the genetic book by some kind of chemical commas, or by a different helicity or configuration. Possibly the "bending" of DNA chains when they fold into the compact "storage" forms might occur most easily at the "page-number" genes, increasing

further the resemblance to a book. Some physical "closure" of this kind, during storage or during reading-periods, might be the simplest way to obtain chemical inaccessibility of the closed pages; but obviously during read-out, at least, the page-numbers or reference headings themselves would still have to be left accessible. What could satisfy these requirements more easily than physical folding, with the page-headings "out"?

And chemically, the page-heading sections of the chain, in any stages when they really do have attached repressor-protein, should be easy to distinguish from the structural sections of the DNA chain. This might be a simple reason for the alternating dark and light nucleic-acid and protein-containing bands of successive "genes" visible in the stretched-out salivary chromosomes of *Drosophila*. [See Stedman and Stedman (1950) and Bloch (1962) on the possible page-heading role of the histones on DNA.] We see that, with this idea of "reading page-headings," the book analogy is fruitful in suggesting links among an especially large number of phenomena and in throwing a different light on several current directions of research.

IX. Sequential Reference

On the book model, our picture of individual cell development comes to be like the reading of a complex instruction manual with forward-references and back-references and appendixes.

In a protozoan after mitosis or in an ovum after fertilization, we may then think of the genetic book as opening automatically, or being opened, to "Page 1." (In diploid cells, there are two books to be opened together and read out together, not always agreeing with each other; but let us omit this added complication here.) All the genetic information from this first operon, page 1, begins to be read out into structural Messenger-RNA, perhaps by RNA-polymerase already present. Soon the regulator-information from this page or from later pages begins to be read out (by this or another enzyme) into repressor-page-headings (reference-slips?) which can diffuse to their operator-sites on later pages. Some of the enzymes or some of the repressors manufactured from this page of instructions may be made only once; others might be made over and over again as long as the page is open, perhaps depending on whether the Messenger-RNA chains carry "cancel" or "repeat" instructions on their ends.

When "Page 1" has finished being read, or perhaps when enough of its enzyme products have accumulated as "inducers" demanding further chemical treatment, "Page 2" may be opened. "Page 1" might then be' closed by the same "inducers" acting as "co-repressors" on it; or it might be closed later by others. At some point, the conditional repressors may begin to be specific for food or poisons in the environment, and "Page 3," "Page 12," and "Page 64'" may be opened by some inducers, not sequentially, but according to functional "heading." And so on, with further page-reading and forward and backward references, or references to special appendixes, up to some point where "enough" growth products have accumulated, the pages are closed or the RNA-polymerase is "turned off," and the DNA-polymerase "copying" begins, in preparation for mitosis.

For the metazoan, the whole book is not finished by the first cell division, but only "Chapter 1." When the book reopens after mitosis, it may reread certain parts of that chapter again, skipping the bits about "reaction to fertilization," but repeating many of the instructions for growth and cell division. But the cytoplasm is no longer the same as it was before, and the accumulated products and inducers in these daughter cells may also begin to open up new pages, the pages of "Chapter 2." And so on, again, to the end of that chapter; and on, through successive chapters in successive divisions.

On this book-reading model, the rule that "ontogeny recapitulates phylogeny" would have to be translated into the statement that many of the early chapters in the genetic books remain almost unchanged for millions of years.

X. Differentiation: Cross-References from Other Books

As each cell begins to have neighbors, so that the chemical gradients produce different concentrations of inducers and co-repressors in the different cells, the various daughter cells in a given generation no longer open to the same chapters at the same time, but to different chapters. In mosaic embryos and similar cases, where the daughter cells seem to lose some potentialities irreversibly from the very first division, each one might have received only a partial copy of the whole genetic book. But it is not necessary to assume this, if only the inducers in the neighboring cells are sufficiently different to start them reading out from different pages, each leading to a

different sequence of developmental steps or page references that never converge; and if there is little exchange of inducers.

Differentiation of pluripotent cells, where each cell's response does depend on neighboring-cell inducers—with one cell giving one set of inducers and co-repressors to another and getting a different set back from it—has a different book-analogy.

The situation now is more like several copies of the anatomy book lying side by side on the table, being read from by several medical students dividing up an assignment. One says, "I'll look up all the information on bone if you look up all the information on blood and you look up all the information on cartilage." Each gives cross-reference slips to the others, so that the students—or the cells—exchange information and look up pages for each other; but the different books are open to different pages and chapters.

Where there is easy diffusion of inducers between the cells, one can see physically at what point specialization must begin. For there is no reason then why any cell should start reading chapters different from the others, or different from what it would have read if it were isolated, until the 8- to 16-cell stage is reached, where some cells begin to be surrounded, and the gradients can begin to produce a chemical difference between the "center" and the "outside" cells. But from that point on, different cells follow different destinies, each referring on from page to page following its particular sequence determined by its history and its neighbors. (Such "center-outside" considerations may be what limit the number of identical twins that can be born fully formed, since the early-division cells must be separated before they have lost their "unsurrounded" totipotency.)

There is obviously no biological reason why any of the somatic pages should ever contain repressor-orders reading: "Close up the open pages and go back to the one-cell stage." The sense of the information-flow from page to page is unidirectional. It may contain instructions for programming some cyclic sequences, but none for programming a reversal. In fact, the pages in differentiated tissue cells are probably always "self-inducing" for much of their chemistry, with inducer-messages from their open operons to themselves that read: "Stay open to these pages."

Such self-stabilizing enzyme mechanisms that stay "on," once "on," have a "flip-flop" character, a "steady-state" kind of "biological memory" that has been discussed by Szilard (1960). "Dedifferentiation" is a common phenomenon in tissue cultures; but before

it can be carried all the way back to the early fertilized-egg stage, we will have to learn how to turn off the "flip-flop." This may require treatment with co-repressors and inducers for earlier chapters; conceivably this might be done fairly simply some day by isolating a tissue cell from its neighbors and then surrounding it with growing cells from progressively earlier developmental stages.

XI. Stored Configurations

A final analogy between genetic information and a book is that both may be transformed part of the time into compact "wrapped-up" or "stored" configurations "in the bookcase." Such configurations, not accessible either to read-out or copying, are found in the packing of DNA into a phage head, and in the condensation of the cellular chromosomes into the short thick forms visible during mitosis. The shrinkage in length in each case is by factors of hundreds of thousands, and the purpose of these forms must be to make a package hard to damage while it is in transit from one "reading-room" to another.

We know surprisingly little about these compact forms, except that the chains do not get mechanically tangled or broken either while they are condensing or while they are being pulled out to their full length again. Perhaps this simply requires every section to slide in the same direction "along its own length," like slippery nylon or like a snake getting out of a knot; because no tangling or knotting is possible with such a motion.

There are several possible configurations for the compact forms. One is a "coil of coils" or a "coil of coils of coils," like the forms used for the tungsten filaments in lamp bulbs. Another is a "transfer-twist" form, with the strands of a one-strand or two-strand helix being pulled out into numerous "lamp-brush" side arms (Platt, 1955). Both these forms are derivable by small perturbations of a double-helix, but they can lead to excessive twisting between the ends unless special conditions are met, or compensatory reverse twists also occur.

Long chains could also be packed into "watch-spring" coils lying on top of each other. But a stiff helix will not bend into a uniform curve as easily as into a group of straight segments separated by sharp bends. (This can be seen by manipulating an extended household steel tape, of the kind that has a quarter-cylindrical spring "set"; because it also bends most readily where a bend has already

started.) Perhaps therefore the DNA will simply tend to fold back and forth on itself in this way. There are also numerous more complicated alternatives, including the possibility that a double-helix might unwind in packing, with its single strands stabilized by quite different chemical attractions in the compact form—attractions which the form itself proves the existence of. [Rich (1962b) has emphasized the possibility of back-and-forth folding, and its analogy to the 100-angstrom-segment folding found experimentally in many long-chain synthetic polymers.]

The related questions of the mechanochemical forces that draw up the chains will be interesting in their own right. Does a wave of drawing-up propagate along the length, or is the drawing-up simultaneous at all points? Is the process as simple as "salting-out" with ions (Rich, 1962b) which could "go" spontaneously with purified DNA in vitro? Or does it involve the cooperation or pulling together of attached protein chains, comparable to the observed shrinking of the attached spindle-protein fibers during mitosis? X-ray and chemical studies and model studies will be necessary. We have begun to understand something about helix-coil transitions, and something about the behavior of chain polyelectrolytes, but the problem of these 1000-fold changes in length could carry us into wholly new aspects of mechanochemistry (Platt, 1961).

Nevertheless we see that the genetic books can be and are stored away in their "bookcase" between readings, however it may be done.

XII. Summary

In summary, the book model gives us a new way of looking at several current research questions. Some are the coding questions of "sense" and "compsense" strands and the direction of read-out in a double-helix chain. Others are the biochemical-genetic questions of "addresses," "page-headings," and operator-genes; of chemical differences between these and the other parts of the genetic chain; of protein-nucleic-acid translator molecules; and of the coding and site of action of embryological inducer molecules. And others are the problems of the diverse physical configurations of the DNA or RNA genetic chains as the various pages are "opened" or "closed" or are shrunk to the compact "storage" forms—configurations that may differ widely, in the large and perhaps in the small, from the long straight, nonreacting, double-helix forms of the conventional picture.

In general, the analogy between genetic information-transfer and a complex instruction-manual would seem to give us a coherent and fairly accurate schema for relating various kinds of phenomena. Both at the teaching level and at the research level, it seems to put into better perspective many of the biophysical and biochemical details of genetic expression in cell development and tissue differentiation.

REFERENCES

Beer, M. (1961). J. Mol. Biol. 3, 263.
Beerman, W. (1959). ("16th Growing Symposium," p. 83. Ronald Press, New York.
Benzer, S. (1959). Proc. Natl. Acad. Sci. U.S. 45, 1607.
Bishop, J., Leahy, J., and Schweet, R. (1960). Proc. Natl. Acad. Sci. U.S. 46, 1030.
Bloch, D. P. (1962). Proc. Natl. Acad. Sci. U.S. 48, 324.
Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). Nature 192, 1227.
Dintzis, H. (1961). Proc. Natl. Acad. Sci. U.S. 47, 247.
Hommes, F. A., van Leeuwen, G., and Zilliken, F. (1962). Biochim. Biophys. Acta 56, 320.
Jacob, F., and Monod, J. (1961). J. Mol. Biol. 3, 318.
Lash, J., Hommes, F. A., and Zilliken, F. (1962). Biochim. Biophys. Acta 56, 313.
Matthaei, J. H., Jones, O. W., Martin, R. G., and Nirenberg, M. W. (1962). Proc. Natl. Acad. Sci. U.S. 48, 666.
Monod, J., and Jacob, F. (1961). Cold Spring Habor Symposia Quant. Biol. 26, 389.
Moscona, A. (1961a). Exptl. Cell Research 22, 455.
Moscona, A. (1961b). Sci. American 205, 142.
Platt, J. R. (1955). Proc. Natl. Acad. Sci. U.S. 41, 181.
Platt, J. R. (1961). J. Theoret. Biol. 1, 342.
Rich, A. (1960). Proc. Natl. Acad. Sci. U.S. 46, 1044.
Rich, A. (1961). "19th Growth Symposium," p. 3. Ronald Press, New York.
Rich, A. (1962a). This volume, p. 103.
Rich, A. (1962b). Proc. Pontifical Acad. Vatican, p. 137.
Speyer, J. F., Lengyel, P., Basilio, C., and Ochoa, S. (1962). Proc. Natl. Acad. Sci. U.S. 48, 441.
Stahl, F. W., in press.
Stedman, E., and Stedman, E. (1950). Nature 166, 780.
Szilard, L. (1960). Proc. Natl. Acad. Sci. U.S. 46, 277.