Central Dogma of Molecular Biology

by
FRANCIS CRICK
MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge CB2 2QH

The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid.

"The central dogma, enunciated by Crick in 1958 and the keystone of molecular biology ever since, is likely to prove a considerable over-simplification."

THIS quotation is taken from the beginning of an unsigned article[1] headed "Central dogma reversed", recounting the very important work of Dr Howard Temin[2] and others[3] showing that an RNA tumour virus can use viral RNA as a template for DNA synthesis. This is not the first time that the idea of the central dogma has been misunderstood, in one way or another. In this article I explain why the term was originally introduced, its true meaning, and state why I think that, properly understood, it is still an idea of fundamental importance.

The central dogma was put forward[4] at a period when much of what we now know in molecular genetics was not established. All we had to work on were certain fragmentary experimental results, themselves often rather uncertain and confused, and a boundless optimism that the basic concepts involved were rather simple and probably much the same in all living things. In such a situation well constructed theories can play a really useful part in stating problems clearly and thus guiding experiment.

The two central concepts which had been produced, originally without any explicit statement of the simplification being introduced, were those of sequential information and of defined alphabets. Neither of these steps was trivial. Because it was abundantly clear by that time that a protein had a well defined three dimensional structure, and that its activity depended crucially on this structure, it was necessary to put the folding-up process on one side, and postulate that, by and large, the polypeptide chain folded itself up. This temporarily reduced the central problem from a three dimensional one to a one dimensional one. It was also necessary to argue that in spite of the miscellaneous list of amino-acids found in proteins (as then given in all biochemical textbooks) some of them, such as phosphoserine, were secondary modifications; and that there was probably a universal set of twenty used throughout nature. In the same way minor modifications to the nucleic acid bases were ignored; uracil in RNA was considered to be informationally analogous to thymine in DNA, thus giving four standard symbols for the components of nucleic acid.

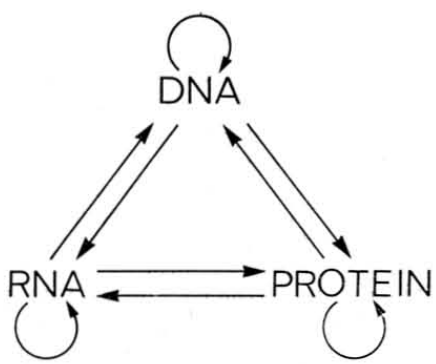

Fig. 1. The arrows show all the possible simple transfers between the three families of polymers. They represent the directional flow of detailed sequence information.

The principal problem could then be stated as the formulation of the general rules for information transfer from one polymer with a defined alphabet to another. This could be compactly represented by the diagram of Fig. 1 (which was actually drawn at that time, though I am not sure that it was ever published) in which all possible simple transfers were represented by arrows. The arrows do not, of course, represent the flow of matter but the directional flow of detailed, residue-by-residue, sequence information from one polymer molecule to another.

Now if all possible transfers commonly occurred it would have been almost impossible to construct useful theories. Nevertheless, such theories were part of our everyday discussions. This was because it was being tacitly assumed that certain transfers could not occur. It occurred to me that it would be wise to state these preconceptions explicitly.

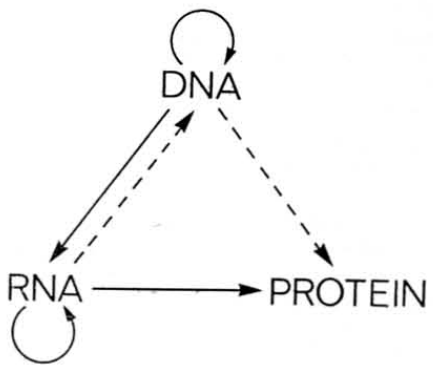

Fig. 2. The arrows show the situation as it seemed in 1958. Solid arrows represent probable transfers, dotted arrows possible transfers. The absent arrows (compare Fig. 1) represent the impossible transfers postulated by the central dogma. They are the three possible arrows starting from protein.

A little analysis showed that the transfer could be divided roughly into three groups. The first group was those for which some evidence, direct or indirect, seemed to exist. These are shown by the solid arrows in Fig. 2. They were:

I (*a*) DNA→DNA
I (*b*) DNA→RNA
I (*c*) RNA→Protein
I (*d*) RNA→RNA

The last of these transfers was presumed to occur because of the existence of RNA viruses.

Next there were two transfers (shown in Fig. 2 as dotted arrows) for which there was neither any experimental evidence nor any strong theoretical requirement. They were

II (*a*) RNA→DNA (see the reference to Temin's work[2])
II (*b*) DNA→Protein

The latter was the transfer postulated by Gamow, from (double stranded) DNA to protein, though by that time his particular theory had been disproved.

The third class consisted of the three transfers the arrows of which have been omitted from Fig. 2. Thes were the transfers:

III (*a*) Protein→Protein
III (*b*) Protein→RNA
III (*c*) Protein→DNA

The general opinion at the time was that class I almost certainly existed, class II was probably rare or absent, and that class III was very unlikely to occur. The decision had to be made, therefore, whether to assume that only class I transfers occurred. There were, however, no overwhelming structural reasons why the transfer in class II should not be impossible. In fact, for all we knew, the replication of all RNA viruses could have gone by way of a DNA intermediate. On the other hand, there were good general reasons against all the three possible transfers in class III. In brief, it was most unlikely, for stereochemical reasons, that protein→protein transfer could be done in the simple way that DNA→DNA transfer was envisaged. The transfer protein→RNA (and the analogous protein→DNA) would have required (back) translation, that is, the transfer from one alphabet to a structurally quite different one. It was realized that forward translation involved very complex machinery. Moreover, it seemed unlikely on general grounds that this machinery could easily work backwards. The only reasonable alternative was that the cell had evolved an entirely separate set of complicated machinery for back translation, and of this there was no trace, and no reason to believe that it might be needed.

I decided, therefore, to play safe, and to state as the basic assumption of the new molecular biology the non-existence of transfers of class III. Because these were all the possible transfers from protein, the central dogma could be stated in the form "once (sequential) information has passed into protein it cannot get out again"[4]. About class II, I decided to remain discreetly silent.

At this stage I must make four points about the formulation of the central dogma which have occasionally produced misunderstandings. (See, for example, Commoner[5]: his error has been pointed out by Fleischman[6] and on more general grounds by Hershey[7].)

(1) It says nothing about what the machinery of transfer is made of, and in particular nothing about errors. (It was assumed that, in general, the accuracy of transfer was high.)

(2) It says nothing about control mechanisms—that is, about the rate at which the processes work.

(3) It was intended to apply only to present-day organisms, and not to events in the remote past, such as the origin of life or the origin of the code.

(4) It is not the same, as is commonly assumed, as the sequence hypothesis, which was clearly distinguished from it in the same article[4]. In particular the sequence hypothesis was a positive statement, saying that the (overall) transfer nucleic acid→protein did exist, whereas the central dogma was a negative statement, saying that transfers from protein did not exist.

In looking back I am struck not only by the brashness which allowed us to venture powerful statements of a very general nature, but also by the rather delicate discrimination used in selecting what statements to make. Time has shown that not everybody appreciated our restraint.

So much for the history of the subject. What of the present? I think it is clear that the old classification, though useful at the time, could be improved, and I suggest that the nine possible transfers be regrouped tentatively into three classes. I propose that these be called general transfers, special transfers and unknown transfers.

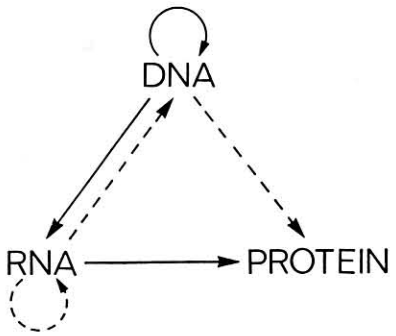

Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

General and Special Transfers

A general transfer is one which can occur in all cells. The obvious cases are

DNA→DNA
DNA→RNA
RNA→Protein

Minor exceptions, such as the mammalian reticulocyte, which probably lacks the first two of these, should not exclude.

A special transfer is one which does not occur in most cells, but may occur in special circumstances. Possible candidates are

RNA→RNA
RNA→DNA
DNA→Protein

At the present time the first two of these have only been shown in certain virus-infected cells. As far as I know there is no evidence for the third except in a special cell-free system containing neomycin[8], though by a trick it could probably be made to happen, using neomycin, in an intact bacterial cell.

Unknown Transfers

These are the three transfers which the central dogma postulates never occur:

Protein→Protein
Protein→DNA
Protein→RNA

Stated in this way it is clear that the special transfers are those about which there is the most uncertainty. It might indeed have "profound implications for molecular biology"[1] if any of these special transfers could be shown to be general, or—if not in all cells—at least to be widely distributed. So far, however, there is no evidence for the first two of these except in a cell infected with an RNA virus. In such a cell the central dogma demands that at least one of the first two special transfers should occur—this statement, incidentally, shows the power of the central dogma in making theoretical predictions. Nor, as I have indicated, is there any good theoretical reason why the transfer RNA→DNA should not sometimes be used. I have never suggested that it cannot occur, nor, as far as I know, have any of my colleagues.

Although the details of the classification proposed here are plausible, our knowledge of molecular biology, even in one cell—let alone for all the organisms in nature—is still far too incomplete to allow us to assert dogmatically that it is correct. (There is, for example, the problem of the chemical nature of the agent of the disease scrapie:

see the articles by Gibbons and Hunter[9] and by Griffith[10]. Nevertheless, we know enough to say that a non-trivial) example showing that the classification was wrong could be an important discovery. It would certainly be of great interest to find a cell (as opposed to a virus) which had RNA as its genetic material and no DNA, or a cell which used single-stranded DNA as messenger rather than RNA. Perhaps the so-called repetitive DNA is produced by an RNA→DNA transfer. Any of these would be of the greatest interest, but they could be accommodated into our thinking without undue strain. On the other hand, the discovery of just one type of present day cell which could carry out any of the three unknown transfers would shake the whole intellectual basis of molecular biology, and it is for this reason that the central dogma is as important today as when it was first proposed.

Received July 8, 1970.

[1] *Nature*, **226**, 1198 (1970).
[2] Temin, H. M., and Mizutani, S., *Nature*, **226**, 1211 (1970). This article contains the references to Dr Temin's earlier work dating back to 1963.
[3] Baltimore, D., *Nature*, **226**, 1209 (1970). See also the brief account of Spiegelman's recent work on page 1202.
[4] Crick, F. H. C., in *Symp. Soc. Exp. Biol., The Biological Replication of Macromolecules*, XII, 138 (1958).
[5] Commoner, B., *Nature*, **220**, 334 (1968).
[6] Fleischman, P., *Nature*, **225**, 30 (1970).
[7] Hershey, A. D., *Nature*, **226**, 697 (1970).
[8] McCarthy, B., and Holland, J. J., *Proc. US Nat. Acad. Sci.*, **54**, 880 (1965).
[9] Gibbons, R. A., and Hunter, G. D., *Nature*, **215**, 1041 (1967).
[10] Griffith, J. S., *Nature*, **215**, 1043 (1967).

Characterization of the Products of RNA-directed DNA Polymerases in Oncogenic RNA Viruses

by
S. SPIEGELMAN
A. BURNY
M. R. DAS
J. KEYDAR
J. SCHLOM
M. TRAVNICEK
K. WATSON

Institute of Cancer Research,
Columbia University, and
College of Physicians and Surgeons,
99 Fort Washington Avenue, New York, NY 10032

Several RNA tumour viruses contain an enzyme that synthesizes a DNA–RNA hybrid using the single stranded viral RNA as template. Hybridization experiments confirm that the DNA strand is complementary to the viral RNA.

TEMIN's DNA provirus hypothesis[1], according to which the replication of the RNA or RNA tumour viruses takes place through a DNA intermediate, explained the following unique features of infections with RNA oncogenic viruses: (*a*) the heritably stable transformation of normal cells induced with these viruses; (*b*) the apparent vertical transmission of high leukaemia frequency in reciprocal crosses between high and low frequency strains of mice[2]; and (*c*) the requirement for DNA synthesis[3] in the early stages of infection.

The hypothesis makes two specific predictions amenable to experimental test. DNA complementary to viral RNA should appear after infection and therefore should be detectable by molecular hybridization. Suggestive but not decisive experiments supporting this prediction have been reported[4,5]. Further, Temin invokes the existence of an enzyme that can carry out a reversal of transcription by catalysing the synthesis of DNA on an RNA template. Evidence for such an enzyme has been presented recently by Baltimore[6] and Temin and Mizutani[7], who found a DNA-polymerizing activity in both avian and murine tumour viruses. The enzyme was detected by the incorporation of tritium-labelled thymidine triphosphate (^{3}H-TTP) into an acid-insoluble product that can be destroyed by deoxyribonuclease. Maximal activity required the presence of all four deoxyriboside triphosphates and magnesium. The fact that the activity is inhibited by ribonuclease implies that the RNA of the virion is necessary for the reaction.

These findings are clearly pregnant with implications for the molecular details of viral oncogenesis. Their potential importance demands quick confirmation and extension, a task the present work undertook to fulfil.

We report here the finding of DNA polymerase activity in all of the seven tumour viruses we have examined and establish by physical and chemical characterization that the product is in fact a DNA heteropolymer. Further, we show that the DNA synthesized is complementary to viral RNA by demonstrating its ability to hybridize specifically with homologous viral RNA. Finally, we find the expected nascent RNA–DNA complexes in the reaction. These have been detected and characterized in glycerol and Cs_2SO_4 gradients and shown to be sensitive to denaturation procedures which disrupt RNA–DNA hybrids.

Preparation of Viruses for Enzyme Test

Rauscher murine leukaemia virus (RLV) was obtained as a ten-fold mouse plasma concentrate. Virus lot RPV-*HL*-67-5 (infectivity titre of 3·9 log spleen weight enlarging units per ml.) prepared from CFW-S mice was used. All procedures following the original thawing of the plasma were conducted at 0°–4° C. Plasma was first clarified at 16,000*g* for 10 min. The resulting supernatant was layered on a 100 per cent glycerol cushion and centrifuged at 95,000*g* for 70 min. The material obtained on and just above the glycerol cushion was then layered over a preformed 25–50 per cent sucrose gradient and centrifuged at 95,000*g* for 3 h. The resulting virus band (1·16 g/cm^3 was diluted in 0·01 M Tris-HCl (*p*H 8·3), 0·1 M NaCl, 0·002 M EDTA buffer (TNE) and recentrifuged for 2 h at 95,000*g*. The resulting pellet was resuspended in TNE and assayed for protein content. A similar procedure served to purify RLV harvested from JLS-V5 tissue culture supernatants grown in our laboratory.