A Conversation about Central Dogma of Molecular Biology

Student: What is the Central Dogma of Molecular Biology?

Teacher: The Central Dogma of Molecular Biology was proposed by Sir Francis Crick in paper he published in 1958 [[1]]. In this paper, Crick discussed a theoretical framework for the mechanisms leading to protein synthesis; for this Crick presented two general principles, which he termed the Sequence Hypothesis and the Central Dogma, though there was scant experimental evidence for either.

Student: OK. So what does the Sequence Hypothesis state?

Teacher: In Crick's own words, the sequence hypothesis "*assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a simple code for the amino acid sequence of a particular protein.*"

Student: Can you explain it in simpler words?

Teacher: There are two parts to this statement. One: the unique identity, or specificity, of a segment of nucleic acids it its sequence. For example, a dinucleotide AT is different from a dinucleotide TA (remember that the sequence of nucleic acids is written in the 5'→3' direction). Two: the sequence of a nucleic acid forms the instruction for a sequence of a protein. At the time when Crick made this statement, rules governing the correspondence between nucleotides in the nucleic acid and amino acids in the protein were not known, but this statement provided a stimulus to uncover these rules – this correspondence between nucleotides and amino acids is now called the Genetic Code. Combining the two parts of this statement, we see that the sequence hypothesis states that DNA/RNA segments of differing sequences code for proteins of differing sequences.

Student: OK. Now, what does the Central Dogma state?

Teacher: Again, in Crick's words, the Central Dogma states that "*that once 'information' has passed into protein it cannot get out again.*" The 'information' in a nucleic acid is its sequence of nucleotides. The 'information' in a protein is its sequence of amino acids. This statement by Crick means that a sequence of amino acids cannot be used to provide instruction for synthesis of either a nucleic acid or a protein. Crick elaborated on this point as follows "*the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible.*"

Student: This sounds clear enough. But I remember reading on Wikipedia[2] that the Central Dogma relates to the '*residue by residue transfer of sequential information*', which sounded quite complicated. Can you explain this in simpler terms? Also, is this different from what Crick had said?

Teacher: Let us first consider the phrase '*residue by residue*'. Residues refer to monomeric units in a polymeric molecule. The monomeric units are deoxynucleotides in DNA, ribonucleotides in RNA, and amino acids in proteins.

Now consider '*transfer of sequential information*'. We know that in this context, 'information' is the sequence of nucleotides or amino acids; to clarify this, Wikipedia has termed it as "*sequential information*". I see how the word '*sequential*' can be confusing, since it might suggest a certain order (sequence) of events rather than the actual composition (sequence) of the polymer. "Sequence-based information" might be a clearer term in this context.

Now comes the most interesting term – "*transfer*". It means that the information (sequence) in one

[1] Crick, F.H.C. (1958): On Protein Synthesis. Symp. Soc. Exp. Biol. XII, 139-163.

[2] http://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

molecule can be copied to create another molecule with the same information (sequence). In other words, one molecule ('mother') acts as a template or guide for the synthesis of another molecule ('daughter') with the same information. This property of reproduction – creation of units of similar type - is characteristic of living beings.

Student: So, how does the '*residue by residue transfer of sequential information*' occur?

Teacher: Two types of activities are central to living beings: reproduction and survival. Instructions are required for both these types of activities, and hence there is a requirement for two types of '*transfer of information*'. The transfer required for reproduction is called replication (also called 'vertical transfer'), wherein information is transferred from a parent generation to a progeny generation. The transfer required for survival can be called 'gene expression', wherein information present within a cell is used within the same cell for carrying out various processes such as respiration that are necessary to meet its immediate requirements.

Student: I understand replication. But what is a "gene" and how is it "expressed"?

Teacher: It is not easy to define a gene since new discoveries are continually changing our concept of a gene [3,4]. For the purpose of our discussion of the Central Dogma of Molecular Biology, we can consider a gene to comprise one or more segments of one or more molecules that need to be copied to carry out a particular function. Continuing with this simple definition of a gene, we can say that gene expression comprises processes involving transfer of sequence-based information present in a gene.

Student: Are there differences between transfer of information in replication and gene expression?

Teacher: Yes, there are three main differences in terms of transfer of information.

First, the purpose of replication is to produce an identical copy of a molecule. Therefore, each residue in one parent molecule is copied to produce one daughter molecule of the same type. Replication is complete when there are two identical molecules – that is, molecules with the same sequence information.

On the other hand, the purpose of gene replication is to meet the immediate needs of a cell. Hence only that segment of a molecule which carries information relevant for a particular requirement is copied.

Second, only one complete copy of a molecule is made at the end of a replication cycle. However, during gene expression, the need might be for some abundant substance, so multiple copies of a particular segment of a molecule may be made. In technical terms, we can say that the stoichiometry or ratio of template: product is 1:1 in replication but is often 1:many during gene expression.

Finally, the end-product of replication is a molecule of the same type as the template molecule, though there might be intermediate molecules of different types. For example, the end-product of replication of DNA is DNA, thought it might occur via an RNA intermediate. But the end-product of gene replication is often a molecule of a type different from the template. In most cells, the template for gene expression is DNA, but the end-product is protein.

Student: OK. Can you tell me now about transfer of sequence-based information during replication?

Teacher: In all cellular organisms, the genetic material is double strand DNA, that is, ds DNA the molecule that determines the identity of a cell. For replication, one molecule of DNA has to be copied to make two identical molecules of DNA.

As you know, the two polynucleotide strands comprising a molecule of ds DNA are complementary to each other. This means that an A residue on one strand is paired with a T residue on the other strand, and a G residue on one strand is paired with a C residue on the other strand. The simplest mechanism to achieve replication is as follows: the two strands comprising a ds DNA

[3] Pearson H (2006) Nature 441: 399-401
[4] Gerstein et al. (2007) Genome Res. 17: 669-681

molecule are separated, and each strand acts as a template for the synthesis of the complementary strand. Synthesis of the complementary strand is done using the Watson-Crick base pairing: A pairs with T, and G pairs with C. In this way, two identical molecules of ds DNA are produced from one molecule of ds DNA.

Some viruses (such as M13 and phiX174) have a single stranded DNA genome. To replicate a ss DNA genome, the DNA is first copied using complementary base pairing to produce a complementary strand, which is then copied to produce copies of genomic DNA.

Replication is catalyzed by an enzyme called DNA polymerase (DNAP). The transfer of information during replication of DNA can be depicted as shown in figure 1.

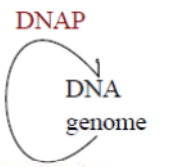

Figure 1: Transfer of information during direct replication of DNA genome

Student: This looks simple. Are there other routes for transfer of sequence-based information for the purpose of replication?

Teacher: Yes, there is another route for replication of DNA genomes, as seen in cauliflower mosaic virus (a plant virus) and hepatitis B virus (an animal viruses), where the DNA genome replicates by a circuitous route.

In this route, genomic DNA is copied to make a RNA copy, which is then copied back to produce the DNA genome. In both these steps, synthesis of the complementary strand is done using the Watson-Crick base pairing: A pairs with T, and G pairs with C.

The first step, synthesis of a RNA copy, is called transcription, and is carried out by an enzyme called RNA polymerase (RNAP). The second step, synthesis of the DNA copy, is called reverse transcription, and is carried out by an enzyme called reverse transcriptase (RT). This can be depicted as shown in figure 2.

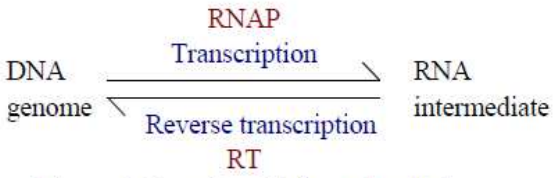

Figure 2: Transfer of information during RT-mediated replication of DNA genome

Student: Why are these steps called 'transcription' and 'reverse transcription'?

Teacher: Well, the four building blocks for DNA are dATP, dGTP, dCTP, and dTTP, whereas the four building blocks for RNA are rATP, rGTP, rCTP, and rUTP. Historically, the use DNA as a template to synthesize RNA was observed first; this process was termed transcription to denote that it involved a change of 'script' of the information from dNTPs to rNTPs.

Later, when the use of RNA as a template to synthesize DNA was observed, the process was termed 'reverse transcription' to note that identities of template and intermediate molecules in transcription was reversed in this process, that is, the template was RNA and the intermediate was DNA.

Student: OK. So life forms with DNA as their genetic material transfer sequence-based information during replication either by DNA → DNA or by DNA → RNA → DNA.

Teacher: Correct.

Student: Are there any other routes for sequence-based information for the purpose of replication?

Teacher: Yes, some viruses have RNA as their genetic material. In viruses like TMV (tobacco mosaic virus; ss RNA genome) and reovirus (ds RNA genome), transfer of information takes from RNA to RNA using RNA intermediates, if necessary. Again, complementary base pairing is used to copy the information. The enzyme catalyzing this process is called RNA-dependent RNA polymerase (RDRP), or simply, replicase. This is depicted in figure 3.

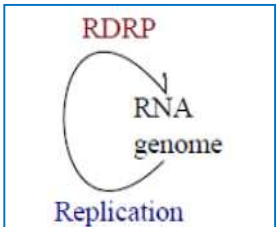

Figure 3: Transfer of information during direct replication of RNA genome

Student: Is there a circuitous way to transfer sequence-based information from a parent RNA genome to a daughter RNA genome?

Teacher: Yes, indeed. In some viruses such as HIV, which belongs to a group called retroviruses, the RNA genome is replicated via a DNA intermediate. You already know the names for the steps involved and the names of enzymes catalyzing these steps. The circuitous route to replicate RNA is depicted in figure 4.

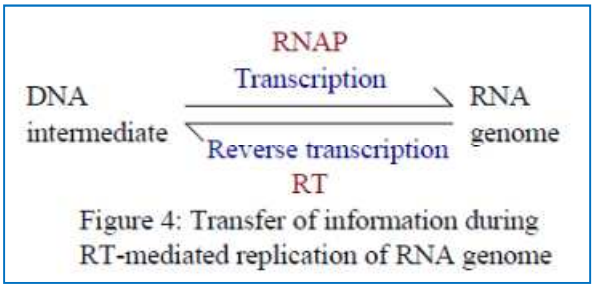

Figure 4: Transfer of information during RT-mediated replication of RNA genome

Student: There seem to be so many different types of polymerases required for transfer of sequence-based information during replication! Is it important to remember all these names?

Teacher: Actually, there are only 4 types of nucleic acid polymerases:

1. DNA-dependent DNA polymerase, or DNAP, which transfers information DNA → DNA.
2. DNA-dependent RNA polymerase, or RNAP, which transfers information DNA → RNA.
3. RNA-dependent DNA polymerase, or RT, which transfers information RNA → DNA.
4. RNA-dependent RNA polymerase, or replicase, which transfers information RNA → RNA.

And no, it is not necessary to know and remember the names of these enzymes to understand the Central Dogma of Molecular Biology. Yet I have mentioned these names to draw your attention to two facts: first, that each route for transfer of sequence-based information requires a unique enzyme. There is no single super-copier type of molecule that can use both DNA and RNA as template and/or produce both DNA and RNA as product. Second, that the transfer of information between nucleic acids requires activity of proteins. But proteins play a non-template role in the process of transfer.

Student: OK. I see now that there are examples to validate Crick's statement that "*the transfer of information from nucleic acid to nucleic acid may be possible*". Are there examples of transfer of sequence-based information from protein to protein during replication? I guess what I am asking is: Are there living beings with protein as their genetic material?

Teacher: No, and yes. What I mean is that you are asking two different questions. Yes, there are some life-forms that have protein as their genetic material. These life forms are called prions – each prion comprises a single molecule of protein. Examples of prions are the agents that cause the "mad cow disease" (the technical term is bovine spongiform encephalopathy; BSE) in cattle and Creutzfeldt–Jakob disease (CJD) in man.

Prions have the same amino acid sequence as normal cellular proteins, but they have a different 3-dimensional structure. But no, they do not exhibit transfer of sequence information from protein to protein.

Student: But if prions are life forms, they must be replicating!

Teacher: Yes, but prions do not replicate by synthesizing new molecules of proteins. Instead, a prion protein interacts with a normal cellular protein of the same sequence and changes its shape to 'convert' it a prion. This is depicted in figure 5. There is no specific term for this process, but we can call it 'conversion' to differentiate it from replication.

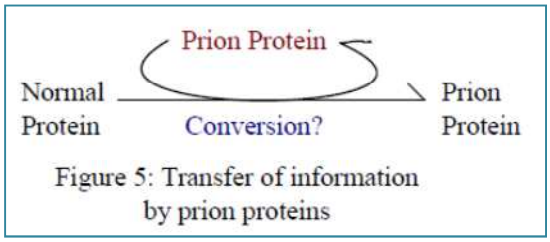

Figure 5: Transfer of information by prion proteins

Student: Interesting! But is this also not a transfer of sequence-based information? After all the 3-dimensional structure of a prion is dependent on its sequence, and a prion 'converts' the structure of only those 'normal' proteins that have the same sequence as the prion protein.

Teacher: It is indeed true that the sequence of a prion protein is a perquisite for its unique structure. But conversion by prions cannot be considered a "transfer of sequence-based information" under the Central Dogma because, though not explicitly stated as such, transfer is related to activity as a template for *de novo* synthesis. Under the Central Dogma a given molecule is considered to "transfer" its information only if a new molecule is synthesized *de novo* (from precursors) using the given molecule as the template.

Student: Why was this point not stated explicitly? If it was not stated, how do you know what qualifies as transfer and what does not?

Teacher: Well, it would definitely have been better if Crick had written a document stating all these points unambiguously. In fact, the Sequence Hypothesis and Central Dogma were presented in just four short paragraphs in a paper titled "On Protein Synthesis", where Crick summarized known experimental evidence and provided speculation or predictions about templates, 'adaptor' molecules, sites, and processes for synthesis of proteins.

A draft version of a previous paper (Oct 1956; link provided in the Wikipedia article on Central Dogma) is also focused on the identity of the template molecule for protein synthesis. It appears that Crick assumed that other scientists in this field (which was just emerging in the 1950s) all understood that the term "transfer" related only to template-based synthesis of molecules with defined sequences.

Student: I see now why I need to learn to describe my observations and ideas in as much detail as possible! Assuming that a reader knows which specific aspect of a property I am referring to may cause confusion. One has to state things explicitly.

Anyway, I think we have covered all modes of 'transfer of information' for replication, which is another term for the process of production of identical molecules. Can you now tell me how transfer of information occurs during gene expression? How does such transfer produce non-identical molecules corresponding to only select segments of a template molecule?

Teacher: As you remember, the aim of gene expression is to carry out metabolic functions in a cell. Most of the metabolic functions in a cell are carried out by proteins.

Each protein is synthesized on ribosomes using a special type of RNA, called messenger RNA or mRNA, as a template. A protein of defined sequence is produced using the mRNA template following the genetic code to establish correspondence between nucleotides on the RNA and amino acids in the polypeptide. This process template-based synthesis of proteins is called translation. This is depicted in figure 6.

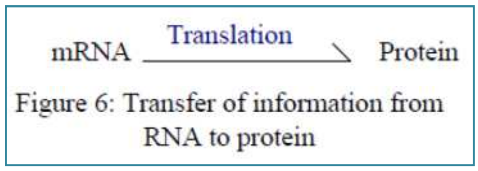

Figure 6: Transfer of information from RNA to protein

Student: Why is this process termed 'translation'?

Teacher: During this step, the sequence-based information contained in mRNA, which is in 'language' of nucleotides, is converted into the 'language' of amino acids. Therefore this step is called translation to denote change of 'language'.

Student: OK. Where does this mRNA come from?

Teacher: In viruses with RNA genomes, the genome itself may act as the mRNA (as for TMV),

or a copy of it may act as mRNA (as for poliovirus). This is depicted in figure 7.

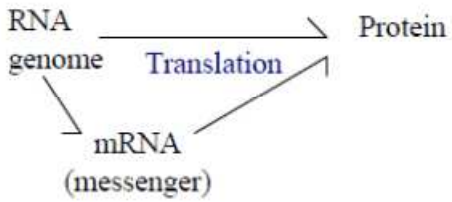

Figure 7: Transfer of information from RNA genome to protein

In viruses and cells with DNA genome, mRNA is produced by transcription of DNA. This is depicted in figure 8.

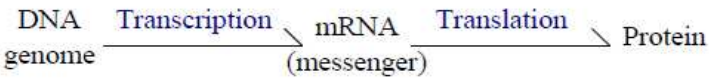

Figure 8: Transfer of information from DNA genome to protein

Student: OK, so the transfers of information in gene expression are DNA → RNA → protein. But, as you said, the entire available template is not copied for gene expression.

Teacher: Right, only some portion of the genomic DNA is copied (by transcription) as mRNA. Sequences in DNA that are not transcribed may act as regulatory regions, and provide binding sites for RNAP and various regulatory factors.

Another point to note is that only the central portion of an mRNA is used as a template for protein synthesis. The 5'- and 3'- untranslated regions (UTRs) contain sequences for binding of ribosomes and other factors that may determine the stability of mRNA.

Student: You have shown me a lot of figures, but none of them look like the one I saw in the textbook.

Teacher: That's right. Most textbooks show the diagram shown in figure 9. This diagram shows the most prevalent routes for transfer of sequence-based information in biological systems – for replication as well as for gene expression.

Student: Can we draw a single comprehensive diagram which shows all known routes of information transfer? Can we show which routes are very common and which are seen only in a few cases? Also, can we show the routes by which transfer of information has not yet been seen?

Teacher: Sure! Such a diagram was first drawn many years ago by Crick [5]. As shown in figure 10, there are nine possible pathways for residue by residue transfer of sequence-based information between DNA, RNA, and protein.

These routes were classified by Crick as general, special, or unknown transfers. General transfers occur commonly, special transfers are seen in few systems, such as viruses and prions, whereas unknown transfers have not yet been shown to exist in any known life form.

Initially, Crick (and George Gamov) thought that a direct transfer of information from DNA to protein was possible, but it has not yet been shown to occur in nature (though it may occur under special conditions *in vitro*). So the arrows that I have classified as general and specific are different from those described by Crick in his paper. But the basic framework of this diagram with the nine routes for information transfer is the same.

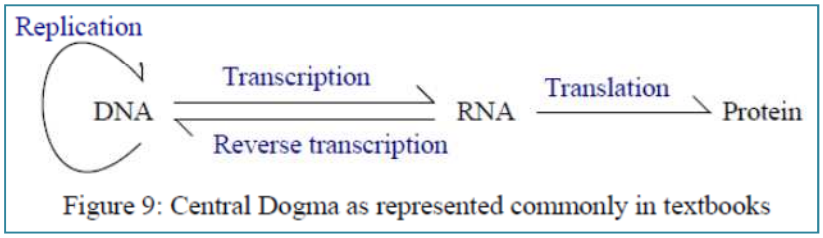

Figure 9: Central Dogma as represented commonly in textbooks

[5] Crick F (1970). "Central dogma of molecular biology". Nature 227 (5258): 561–3.

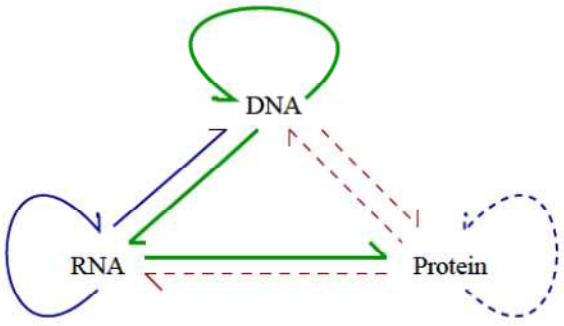

Figure 10: Routes for transfer in genetic information. Thick green arrows show routes for general transfer, thin blue arrows routes for restricted transfer, and red dotted arrows show prohibited routes (not yet shown to occur). Protein to protein transfer is shown as a blue but dotted arrow since it does not involve *de novo* synthesis of proteins, only a change in conformation of pre-exisiting proteins.

Student: OK – I now know all routes for residue to residue transfer of information in life forms. But why did Crick call it the "Central Dogma of Molecular Biology"?

Teacher: In his autobiography '*What Mad Pursuit*', Crick explained why he used the word 'dogma': "*I called this idea the central dogma, for two reasons, I suspect. I had already used the obvious word hypothesis in the sequence hypothesis, and in addition I wanted to suggest that this new assumption was more central and more powerful*.'

But many people were confused by this word. As Crick wrote in his autobiography:"*As it turned out, the use of the word dogma caused almost more trouble than it was worth.... Many years later Jacques Monod pointed out to me that I did not appear to understand the correct use of the word dogma, which is a belief that cannot be doubted. I did apprehend this in a vague sort of way but since I thought that all religious beliefs were without foundation, I used the word the way I myself thought about it, not as most of the world does, and simply applied it to a grand hypothesis that, however plausible, had little direct experimental suppor*t."

Similarly, Horace Freeland Judson wrote in the book '*The Eighth Day of Creation*' that: "*My mind was, that a dogma was an idea for which there was no reasonable evidence. You see?!" And Crick gave a roar of delight. "I just didn't know what dogma meant. And I could just as well have called it the 'Central Hypothesis,' or — you know. Which is what I meant to say. Dogma was just a catch phrase."*

Student: So Crick used the word 'dogma' as a catchy term to denote a powerful type of hypothesis, and not in the actual sense of the word, which means a tenet that cannot be doubted or questioned. And many other scientists also thought like Crick that the word dogma had been used to draw attention to the fact that there was no evidence (available at that time) for the proposed routes for transfer of information. This story goes to show that a scientist should have a good vocabulary and know the correct and precise meaning of words they use. I see once again the need for clear concise text!

Teacher: True. But a mistake in choice of words, or in failing to explicitly define what he meant by 'transfer' does not detract from the value of the idea proposed by Crick.

When reverse transcription was discovered, various scientists made statements to the effect that the Central Dogma had been overturned, because they believed that the dogma simply stated, as put by Marshall Nirenberg, "DNA makes RNA makes protein" - the scheme shown in figure 8.

What we need to learn is that like all hypotheses in science, the "Central Dogma" is a dynamic concept. It can be modified and refined, or rejected outright, to take into account observations made by scientists after it was proposed. And while there was little or no experimental evidence to support the routes of information transfer in the Central Dogma at the time when Crick proposed it, there is plenty of evidence now to let us differentiate between the general, special, and unknown pathways for 'residue by residue transfer of sequential information'.

Student: I see. So might it be a good idea to retire the controversial word "Dogma" and replace it with a word like "Fact" or "Reality"? Perhaps we could call it "Central Facts of Molecular Biology"?

Teacher: Why not, indeed?

Student: Do most of the scientists accept the Central Dogma as the Central Dogma? I mean, are there people who challenge or question the statements which Crick calls the Central Dogma?

Teacher: Yes, most scientists agree that transfer of sequence-based information cannot occur with protein as a template. But many people have criticized the Central Dogma as being incomplete. Crick himself has noted these points in his 1970 paper. He identified four major criticisms of the Dogma by his peers:

1. The Dogma does not mention the machinery required for the transfers or the accuracy of the processes.
2. The Dogma does not mention processes behind the rate and regulation of information transfer.
3. The Dogma describes the routes of information transfer in modern day organisms and does not explain how evolution might have occurred, especially of the genetic code.
4. The Dogma is the same as the Sequence Hypothesis.

Student: So what was Crick's response to these comments?

Teacher: Crick may be one of the masters of understatement. He responded by saying "*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 which statements to make. Time has shown that not everybody appreciated our restraint.*"

So, you see, some people might think Crick said too much in declaring that proteins cannot be used as templates without any evidence to back his statements, while others might think he did not say enough about factors and mechanisms affecting transfer of information (or lack thereof). We can speculate about the reasons for his silence – among others, it could be absence of evidence, lack of interest in that aspect, or a desire to focus only on certain points and a plan to write about other points later. But our focus today is on what he did say, and not on what he did not say.

Student: So true! Looking at his restrained response to critics of Central Dogma, I am reminded of the last sentence in his paper [[6]] describing the double helix model of DNA "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*." And this copying mechanism is what makes transfer of information possible for both replication and gene expression.

But what happens if there are errors in copying a template – during replication, transcription, reverse transcription, or translation? Do errors happen often? If so, does it violate the Central Dogma?

Teacher: Well, you know that the Central Dogma did not make any statement about the extent of accuracy in copying a template, or about the consequences of errors. We now know that each polymerase has an intrinsic limit of fidelity; in general, DNAP makes fewer errors that RT, RNAP, or replicase. Presence of damaged or modified bases, or other chemicals in the cell can increase the error rates in replication, transcription, and reverse translation.

Error in replication produced daughter molecules that are different (or mutated) compared to the parent molecules; these errors are copied during replication to produce the next daughter molecule. Similarly, errors in gene expression can result in metabolic havoc. But that does not mean that the routes for transfer of sequence-based information are inoperative! The Dogma still holds.

Student: I have another question: what does masking of DNA by proteins to make certain

[6] Watson J and Crick F. (1953). *Nature* **171**: 737-738.

sequences unavailable for copying mean to the Central Dogma? In this case, proteins are controlling the transfer of information, so can one claim that information is transferred from DNA alone?

Teacher: Blocking transfer of information under certain conditions does not mean that such a transfer is not possible theoretically. The Central Dogma tells us what is possible, not if it will actually happen. We can compare it to thermodynamic equations that tell us which reactions are favored, but not whether they will actually happen, and at what rate, under a particular set of conditions.

In fact, in his paper outlining the Central Dogma, Crick says that the Sequence Hypothesis provides a view uniting "*the central biochemical importance of proteins and the dominating biological role of genes*", which shows that he was aware of the key roles played by proteins in all cellular processes.

We saw that one of the criticisms of the Dogma is that it does not explain if and how regulation occurs – of replication or gene expression. It is now over 50 years since the Central Dogma was proposed. But no 'central' or 'universal' hypothesis has been put forward to explain this aspect of biology. So one can hardly fault Crick for not discussing the aspect of regulation in his Dogma. Perhaps it was a lucky omission, or perhaps he had the foresight to know that the diversity of regulatory processes in biology could not be described by a few simple principles.

So let me ask you one question now. Do you think that the Sequence Hypothesis and the Central Dogma are two names for the same idea?

Student: Well, no. It is true that these two ideas are related, but they are distinct.

Simply put, the Sequence Hypothesis states that the sequence of a nucleic acid is its identity as well as the code for replication and gene expression. The unstated part is that the sequence of a protein is its identity, but not a code for replication and gene expression.

The Central Dogma states that DNA and RNA can act as templates for the synthesis of DNA, RNA, and proteins. Proteins cannot act as templates for the synthesis of DNA, RNA, or proteins.

But I can understand why some people might be confused on this point. What could have made Crick propose the Central Dogma in two mutually contradictory forms? In his 1958 paper, he states clearly transfer of sequence information "*from protein to protein, or from protein to nucleic acid is impossible*", yet in his 1970 paper, he talks of nine pathways (some general, some special, and some unknown) for information transfer.

My other, and more fundamental questions is this: In 1958, when there was no experimental evidence to rule out transfer of information from proteins to nucleic acids or proteins, why did Crick not make a simple null hypothesis-like statement: "*All modes of transfers are possible unless proved otherwise*"?

Teacher: A hypothesis stating "*all modes are possible/equally likely unless proved otherwise*" or the converse "*no modes are possible unless proved otherwise*" may sound like a null hypothesis for inferential statistical testing (where the null hypothesis, H_0, typically states that there is no difference between the 'control' and 'treated' sample with respect to the test statistic). However, such observations are self-evident and of little actual value in the context of describing possible biological processes.

Obviously, one can argue that nothing should be ruled out until proper proof is provided. But it is a valid exercise in science to predict that certain events are likely or unlikely to happen, based on some preliminary knowledge and extrapolation. If one's predictions come true, they get credit; if not, one should be prepared for modification or dismissal of the hypothesis.

When Crick proposed the Central Dogma, he noted that he needed some framework to think about the complex problem of protein synthesis. Reconstructing what was known at that time, we can surmise that Crick knew that DNA was the genetic material (experiments of Avery's group). He also knew that viruses like TMV had RNA as

genetic material. He knew that genes and polypeptides were collinear (the prevailing notion of a gene at that period was "one gene produces one polypeptide").

So it stood to reason that genomic DNA and/or RNA carried instructions for synthesis of proteins. If the sequence information in proteins could be used to synthesize a corresponding nucleic acid, there would be no need to have DNA/RNA as genetic material; biological systems comprised entirely of proteins would be the most economical and efficient models of replication. This was the logic behind why Crick ruled out proteins as a source of sequence information.

As stated in his 1970 paper: "*In brief, it was most unlikely, for steriochemical resons, that protein → protein transfer could be done in the simple way that DNA → DNA transfer was envisaged.*" He goes on to say that the machinery required for translation was very complex, and unlikely to work 'backwards' to use a protein template to produce RNA or DNA. This would mean that an independent but equally complex machinery would be required for such "back translation", and since "*of this there was no trace, and no reason to believe that it might be needed*", Crick decided to rule out information transfer from proteins.

Crick has stated that found that the statements made in the Sequence Hypothesis and the Central Dogma stimulated him to think about the process of protein synthesis and make some predictions about how the process would operate. Two very important predictions – which were later shown to be true – came out of this exercise:

1. all ribosomes had the same composition and that they were not simply complexes of the particular protein being translated, and
2. there were two types of RNA (the message mRNA and the adaptor tRNA) in systems with DNA genome (especially eukaryotes, where DNA is in the nucleus and ribosomes in the cytoplasm).

Statements of the Sequence Hypothesis and Central Dogma helped Crick to predict the nature of the genetic code. These predictions helped other scientists in designing experiments to understand the role of mRNA, tRNA, and the genetic code.

So, now I will ask you a question: Why did Crick rule out proteins as a source of sequence information? Was it because he did not understand how to properly formulate a hypothesis, or was he simply making a random (or perhaps an educated) guess, or was it a culmination of thought processes based on several preliminary but disjointed pieces of evidence?

Student: Before I answer this question, I want to ask one final question. As of now, there are no known cases that violate the Central Dogma, which states that "that once 'information' has passed into protein it cannot get out again." Does this mean prove the Central Dogma is correct?

Teacher: Of course not! Absence of proof is not proof of absence. Just because we have not yet found any biological system where the sequence of a protein is necessary and sufficient to direct the synthesis of an identical sequence of amino acids, or a corresponding sequence of nucleotides, does not mean that such a biological system does not or cannot exist. Absence of proof cannot be taken as the basis to validate a hypothesis.

Student: OK. So I see that Crick has made certain statements which he termed the Central Dogma. All experimental evidence supports his statements, and so far nothing has been found to invalidate his statements. In the absence of any other competing hypothesis that makes an alternate but equally likely statement regarding the flow of genetic information in biological systems, one cannot dismiss the Central Dogma as wrong. This is not the same as saying that it is right.

As for why Crick ruled out transfer of sequence from a protein to another protein or DNA, his writings point to a logical conjecture on his part, so I cannot say that he made a lucky guess, or that he did not know how to state a hypothesis.

As for why Crick called his statement a Dogma, I see it as more an honest error in comprehending the exact meaning of the word 'dogma' than a declaration that his statement cannot be doubted and that it should be accepted unquestioningly as a matter of faith.

Teacher: Here I would like to quote a few more sentences from Crick's paper of 1970. After the discovery of reverse transcription, there was some debate about the validity and relevance of the Central Dogma.

"*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 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.*"

It is remarkable that Crick proposed that repetitive DNA may be generated via RNA templates. Today we know of at least two such instances: telomeres and retrotransposons.

Student: What? Do human cells also contain RT? I thought RT was restricted to certain types of viruses. And where does the RNA template for reverse transcription come from?

Teacher: Actually, RNA templates are integral for one mode of replication of DNA in linear chromosomes, right from yeast to human cells. In these cells, each end of the linear chromosome – called a telomere - comprises hundreds of tandem repeats of a simple hexanucleotide sequence. While the body of the linear chromosome is replicated by DNAP using a parental strand as the template, telomeres are added by a special enzyme called telomerase.

Each molecule of telomerase carries a short RNA segment is an integral component; this RNA is used as a template to add the telomeric repeat sequences at the ends of linear chromosome. Note that the entire chromosome is not copied into a RNA intermediate.

Another interesting point about telomerase is that cells with very low levels of telomerase do not divide; they age and eventually die. On the other hand, rapidly dividing cells, such as cancer cells, have high levels of telomerase. But remember that while telomerase may be a factor that influences the decision of whether to replicate genomic DNA in a given cell, telomerase does not affect the sequence-based information present in the DNA.

Student: And what is the story with retrotransposons?

Teacher: Well, transposons are segments of self-replicating DNA; these segments replicate and insert themselves in various seemingly random locations in a chromosome. Transposons have been found in almost all genomes studied so far, from bacteria to mammals.

Simple transposons replicate by DNA → DNA transfer, while retrotransposons, also called retroposons, replicate by DNA → RNA → DNA transfer. The RT required for the RNA → DNA step in the replication of a retroposon is encoded by the retroposon itself. Retroposons have sequence similarity to the RT-containing retroviruses.

Student: I see. So transfer of sequence-based information in telomeres and retrotransposons can be easily accommodated within the Central Dogma. And unless we find a life-form capable of synthesizing a nucleic acid or a protein using only the sequence information contained in a protein, the Central Dogma is sufficient to describe the information-management systems in biological systems.

I have learned so many things today. Other than the various routes for transfer of sequence-based information in biological systems, I learned the purpose for each route.

I also learned that while it is important for an author to use correct words when describing new ideas, the reader too should also try to see the idea conveyed by the words: focusing on individual trees may make us miss the forest! An idea is more than the sum of the words used to convey it.

Thank you teacher! This was an interesting conversation about the Central Dogma.