

# Molecular Biology: The Central Dogma

Patricia J Pukkila, *University of North Carolina at Chapel Hill, North Carolina, USA*

Although proteins are essential components of all aspects of cell structure and function, they are not sufficient for their own propagation. Instead, the information necessary to specify each individual protein is stored in nucleic acids. The 'central dogma' of molecular biology states that sequence information can be transferred among nucleic acids, and from nucleic acids to proteins, but sequence information cannot be transferred among proteins, or from proteins to nucleic acids.

## Introduction

In the twentieth century, the mechanisms governing the specification and transmission of genetic traits were understood for the first time. Living beings are composed of cells, which in turn are composed of many types of chemicals. We are particularly concerned with large 'macromolecules' that enable cellular function and propagation. It is first essential to realize that these complex macromolecules with a bewildering variety of shapes and sizes are composed of simpler subunits. For example, nucleic acids (either DNA or RNA) are made up of four types of subunits called nucleotides, while proteins are composed of 20 types of subunits – the amino acids. Second, a complete description of a nucleic acid or protein macromolecule (complete in that it serves to distinguish one macromolecule from any other) can be obtained by simply determining the type of subunit at each position in the linear chain of subunits. We speak of the particular DNA, RNA or protein sequence because these macromolecules are unbranched linear polymers of subunits joined by just one kind of linkage. Finally, some macromolecules serve to store information necessary to specify the formation of other macromolecules, while others depend on those repositories for their own creation. Given that there is specialization among classes of macromolecules, it is necessary to consider how information is conveyed between these classes. Thus the central issues of molecular biology are to identify the macromolecules that carry information, and to determine how that information is used to make other necessary macromolecules.

## The Concept of Information Flow

The DNA macromolecule is an elegant structure that accommodates the need to store and transmit genetic information. A chromosomal DNA molecule is, in fact, two single chains of nucleotide subunits that are comple-

mentary to each other (**Figure 1a**). Thus there is present both the primary information, and the template necessary to make a new copy of that information within this chemically inert macromolecule. The redundancy helps to ensure that essential information is not lost, because if one copy is damaged, the other can serve as a template for repair. The dangers inherent in transmitting genetic information are also reduced. If only a single chain were present, essential information might be lost either when the necessary template was assembled, or subsequently when the copies were produced.

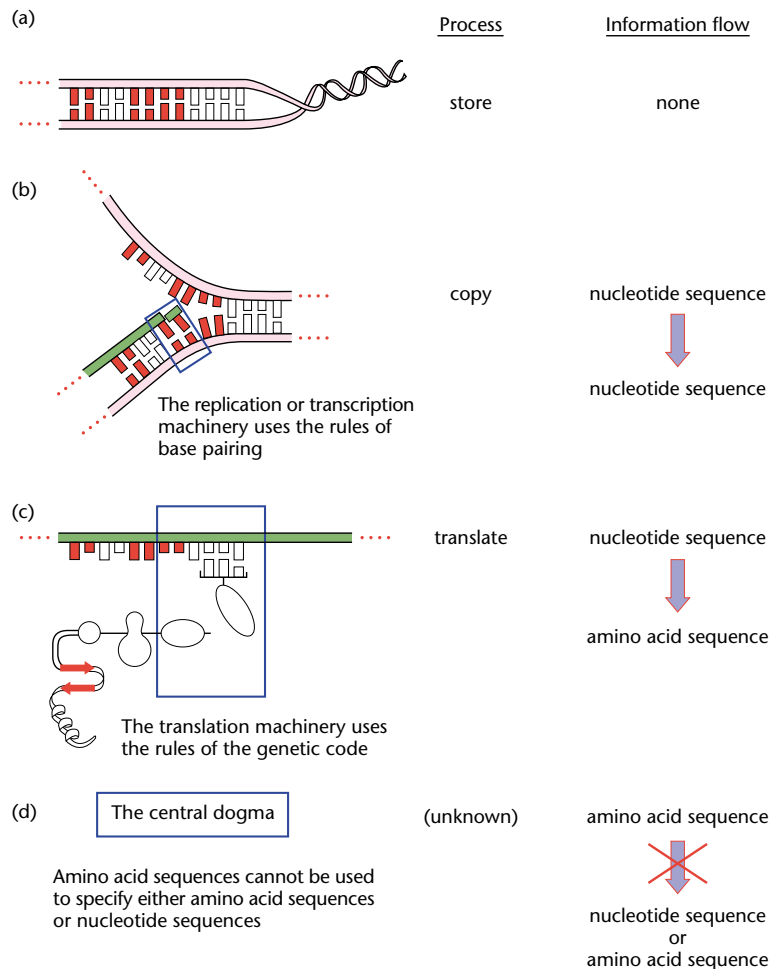
The chemical stability of DNA, while advantageous in its role in data storage, must be overcome if the information present is to be used by the cell and transmitted to other cells. The stability of DNA is altered by mechanisms that temporarily separate the two chains, and then either copy particular portions of the information (so that it can be used) or copy each chain of the entire chromosome (thereby producing two chromosomes in preparation for cell division). These processes of chain separation, preparing working copies of selected portions (transcription), and making copies of the entire chromosome (replication) are carried out by macromolecular 'machines' (**Figure 1b**). Both transcription and replication use the rules of base pairing between complementary nucleotides. However, during transcription, the macromolecular machine produces a chemically less stable copy (so that the genetic information present can be used selectively at different times or by different cell types in the organism). This more labile nucleic acid has a slightly different sugar molecule in each nucleotide subunit, and is called messenger RNA.

Soon after it was discovered that DNA was a linear polymer, it was discovered that proteins are also linear polymers. Since there are only four types of nucleotide subunits in DNA or RNA but 20 types of amino acids, the information transfer between DNA and nascent protein must be indirect. Amazingly enough, one of the simplest possibilities, that a group of three nucleotides is sufficient

## Introductory article

### Article Contents

- Introduction
- The Concept of Information Flow
- The Central Dogma in its Historical Context
- The Reverse Transcriptase Ripple
- Refuting Lamarckism and Interpreting 'Adaptive' Mutation
- Summary



**Figure 1** Information transfer between DNA, RNA and protein macromolecules. (a) Complementary base pairing in a short segment of DNA. The four types of nucleotides are represented by the bars (solid or open, long or short), and each nucleotide can pair with only one complement. (b) Information transfer between DNA (drawn as in part a) and RNA (green). RNA can also be used to specify DNA by an analogous mechanism. (c) Information transfer from RNA to protein. The 20 types of amino acids (the four oval shapes represent four of these) are linked according to the sequence of nucleotides in the messenger RNA, which in turn was copied from the gene in part b. The next amino acid which will be added is shown with its adaptor transfer RNA (tRNA) attached. (d) The central dogma of molecular biology specifies the forbidden information transfers.

to specify a particular amino acid, and that the sequence of these nucleotide triplets along the relevant portion of the DNA corresponds to the amino acid sequence in the protein, proved to be correct (Figure 1c). Nucleic acid ‘adaptors’ (transfer RNAs) contain nucleotide triplets complementary to those in the messenger RNA. Each adaptor must be linked to the appropriate amino acid according to the rules of the genetic code. Conventionally, the word translation refers to the process of RNA-directed protein synthesis, although the macromolecular machines that correctly link each adaptor to the corresponding amino acid are technically responsible for ‘translation’ of the code.

Considerable evidence has accumulated in support of the central mechanism of information flow from nucleic

acid to proteins. A region of the chromosome that harbours information necessary to specify the sequence of amino acids in a particular protein is called a gene (or locus, which means ‘place’). We speak of the gene as ‘coding for’ a particular protein. The information necessary to specify several kinds of proteins that must cooperate to transcribe and replicate DNA is found in the DNA itself, along with information needed to specify all the other macromolecular machines in the cell. Thus there is the necessity for information to flow from DNA (the storage molecule) to other DNA, to RNA (the working copies), and finally to protein (the end product). It is also clear that information might be expected to flow among nucleic acids, since these can (and do) serve as templates for their own synthesis. Both nucleic acids and

proteins are linear polymers, and thus it is conceivable that information in protein molecules could be used to specify the information in nucleic acid molecules rather than just being harnessed as part of the macromolecular machines that copy the nucleic acids. The fact that information does not flow from protein to nucleic acid was first articulated by Francis Crick, and he offered this as the 'central dogma' around which a working model of the gene could be constructed (**Figure 1d**). Since nucleic acids have one form of specificity (due to their ability to engage in complementary base pairing), while proteins have another, a complex machinery is required to transfer information from nucleic acids to proteins. In fact, there is no evidence of any portions of the machinery that would be required to utilize the linear information in a protein and 'back-translate' so as to produce a nucleic acid. Thus the central dogma still has predictive power 40 years after it was proposed.

## The Central Dogma in its Historical Context

Horace Freeland Judson in his book *The Eighth Day of Creation* equates Crick's formulation of the central dogma with Einstein's articulation of  $e = mc^2$ . Crick was planning to address the Society for Experimental Biology at their symposium in September 1957, and chose to summarize what was known about protein synthesis to enable the general reader to understand the issues. As he said to Judson, when a scientist writes a review article, by putting words on the page he or she reveals assumptions that can then be examined. The resulting article makes the difference between conclusions based on evidence and theories for which there was no evidence quite clear. Crick was concerned with two hypotheses: one concerning the importance of considering the gene as a sequence of nucleotides, and the second concerning the differences between nucleic acid and protein specificities. The first he simply called the sequence hypothesis, but the second seemed to require a more exalted status, so he entitled it 'the central dogma'. While dogma can be assumed to be established opinion backed by authority, Crick had in mind the meaning of dogma as a point of view that is presented authoritatively without any proof. In the 1958 paper based on his address, he challenges the reader to construct a useful model of protein synthesis without resorting to the two hypotheses. Indeed, such models are impossible.

What is remarkable about the 1958 paper, and what makes it a pleasure to read today, is the clarity of the connections between the limited number of observations presented. The 'central and direct' role of proteins was acknowledged such that there would be 'little point' in genes doing anything except directing their synthesis. It was known that proteins were composed of amino acids,

and that proteins had complex three-dimensional shapes that are crucial for their proper function. However, the essential clues to their synthesis lay not in their complexity but in their simplicity. Despite the variety of amino acids, Crick emphasized that only 20 were fundamental, and that most of these occurred in all proteins. A protein could be specified as a particular linear array of amino acids. It had recently been shown that a mutant form (allele) of the gene that specified haemoglobin (the allele conferring the sickle cell trait) resulted in a single amino acid substitution. Thus if the action of one gene was clearly to specify a particular sequence alteration, it was logical to at least contemplate that the function of all genes might be to specify particular sequences. This set aside the difficult problem of how the protein folded so as to adopt its essential shape. Crick postulated that the protein might fold up on its own, as it was synthesized. Having proposed a possible mechanism for folding that did not involve the direct action of the gene, Crick did not feel compelled to consider if this was in fact how proteins did fold. By its very nature, a scientific conclusion always has limits. Since he defined the relevant limits at the outset, Crick was free to focus on solving one problem at a time. Thus the sequence hypothesis was articulated.

## The Reverse Transcriptase Ripple

In 1970, two laboratories demonstrated biochemically that RNA could be used as a template for DNA. Howard Temin had argued that the RNA of some viruses must be copied into DNA in certain cell types in order to explain the 'transformation' of normal cells to cancer cells by these viruses. Detecting the enzyme responsible for this RNA-dependent DNA synthesis was an important piece of evidence that convinced many former sceptics that his theory was valid. For many, the new results seemed consistent with the predictions of the central dogma, but for others, these findings were seen as a challenge to it. One writer went so far as to suggest that existence of 'inverted transcription' (copying RNA to make DNA) meant that the entire central dogma needed to be re-examined. In retrospect it does seem a bit strange that this (anonymous) writer did not bother to reread the original formulation before calling for it to be discarded! Crick responded to the challenge and a few weeks later published an expanded view of the central dogma. He reiterated that information transfer from RNA to RNA, from RNA to DNA, or perhaps even from DNA to protein (directly, without an RNA intermediate) were all 'special transfers' which might occur in certain cell types. As such, they were perfectly consistent both with the sequence hypothesis and with the flow of information among nucleic acids or from nucleic acids to proteins. Thus the newly discovered 'reverse transcriptases', as the RNA-dependent DNA polymerases

are now called, did not disturb the central dogma, despite the initial 'ripple'. However, Crick also pointed to three 'unknown transfers', from protein to either protein, RNA or DNA. Such transfers, if shown to exist, would require a radical reformulation of molecular principles. If the gene products (proteins) could alter genes (DNA sequences), the way would be paved for the inheritance of acquired characteristics. In 1970, Crick felt that the available evidence was still insufficient to conclude that the central dogma was certain to be correct, although he maintained that it was likely to remain useful. More recently, it was found that under certain circumstances, the DNA sequences of an organism became altered in what appeared to be a directed way in response to environmental stimuli. In addition to expanding our understanding of the origin of mutations, the experiments raised the possibility that 'advantageous' proteins might cause 'advantageous' mutations. Thus analysis of the origin of 'adaptive' mutations is crucial to evaluating the validity of the central dogma, since these raised the possibility that information might flow from proteins to nucleic acids.

## Refuting Lamarckism and Interpreting 'Adaptive' Mutation

In 1988, John Cairns, Julie Overbaugh and Stephan Miller published a provocative paper in *Nature* entitled 'The origin of mutants'. Until their study, it was generally accepted that alterations in DNA (mutations) occurred randomly, as a consequence of the repair of chemical damage or due to errors arising during replication. These variants might be transmitted to progeny, and thus individuals in large populations would be expected to harbour slight differences in DNA sequences. If the environment changed dramatically, individuals with alterations that were advantageous under the new circumstances would be expected to thrive, and might come to outnumber the individuals with the 'nonmutant' alleles. This paradigm had been established in the 1940s by elegant experiments of Salvador Luria and Max Delbrück using bacterial populations. Normal bacteria will be killed if exposed to a particular bacterial virus. However, in a population of bacteria, mutations can be acquired that have the effect of protecting the host from being killed by the virus. Luria and Delbrück showed that exposing the bacteria to the virus did not increase the likelihood that the bacteria would acquire the beneficial mutation. Instead, the virus 'selected' the bacteria with the protective mutation (they survived), and instantly killed the rest.

Cairns and his colleagues accepted this demonstration that some mutations arise spontaneously, but they wondered if others might arise in a 'directed' fashion. They decided to examine a bacterial population in which an essential nutrient could not be used by the 'normal'

bacteria (no cells were being killed – they simply could not grow or divide). They found that some of the mutations that permitted the cells to use the nutrient occurred after the cells were exposed to that nutrient! Furthermore, mutations that would not assist in metabolizing the rate-limiting nutrient did not appear to accumulate in the experiments. This was a serious challenge to the central dogma. It appeared that mutations were not strictly 'random', but could in fact be 'directed'. The environment appeared to be influencing the genotype so that advantageous traits were being acquired and transmitted to progeny. The authors postulated that errors in transcribing the DNA into messenger RNA might result in a variety of proteins. Production of a favoured protein might trigger a reverse transcriptase complex to produce a DNA copy of the variant RNA, and the genome could be altered by established recombination mechanisms. Thus, information could, in theory, flow from protein to DNA without the requirement for the molecular 'back-translation' machinery.

The paper unleashed a flood of commentary and further experimentation. Appropriately, Cairns himself (with Patricia Foster) provided an elegant disproof of their 'reverse transcriptase' model by showing that some of the advantageous revertants arose in other components of the translation machinery, and thus could not have arisen by reverse transcription of the 'favourable' messenger RNA. In fact, all experiments designed to ask if the mutations were in fact 'directed' revealed that instead, they arose in a nondirected way, although the net result was that they were advantageous to the cell under the particular circumstances (thus, they were 'adaptive'). In looking back over the controversy, the flurry of papers and published arguments underscores the wisdom of using theories to selectively ignore extraneous complexity. Some authors behaved as though they would discount the conclusions of a well-constructed series of experiments if the mechanism underlying the surprising observations was not yet understood. Other authors proposed general models to predict how the phenomenon might work, which freed them to test parts of the models. The step-by-step approach was again productive: it now appears that a mutagenic process involving breaks in DNA and recombinational repair of the breaks with error-prone DNA synthesis occur in a small number of nondividing cells. If an ensuing 'randomly' generated mutation permits the cell to grow and divide, the new genotype will be selected and those individuals will quickly outnumber the others in the population (including those with newly arising mutations that are not immediately beneficial, explaining why these were hard to detect in the initial experiments).

## Summary

The central dogma of molecular biology predicts that a particular sequence of amino acids (a protein) cannot be used to specify or even alter a particular sequence of nucleotides (a gene). Instead, information flows from nucleic acids to proteins, in that an elaborate machinery exists to 'translate' the nucleic acid 'alphabet' to the amino acid 'alphabet' according to the rules of the genetic code. Cells exhibit no trace of a 'back-translation' machinery, and organisms can transmit only their genes to their offspring. Even though the genetic material is not entirely constant, advantageous mutations do not arise in a directed manner. The predictions of the central dogma have withstood every challenge, and are likely to remain as the central organizing principles of molecular biology.

## Further Reading

- Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumor viruses. *Nature* **226**: 1209–1211.
- Cairns J, Overbaugh J and Miller S (1988) The origin of mutants. *Nature* **335**: 142–145.
- Crick FHC (1958) On protein synthesis. *Society for Experimental Biology Symposia* **12**: 138–163.
- Crick FHC (1970) Central dogma of molecular biology. *Nature* **227**: 561–563.
- Foster PL and Cairns J (1992) Mechanisms of directed mutation. *Genetics* **131**: 783–789.
- Judson HF (1979) *The Eighth Day of Creation*. New York: Simon and Schuster.
- Luria SE and Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.
- Temin HM and Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**: 1211–1213.
- Rosenberg SM (1997) Mutation for survival. *Current Opinion in Genetics and Development* **7**: 829–834.
- Stahl FW (1992) Unicorns revisited. *Genetics* **132**: 865–867.