Unit 4: Central dogma and genetic code (2 lectures)
Key experiments establishing-the central dogma (adaptor hypothesis and discovery of mRNA template), Genetic code: salient features and deciphering (triplete binding assay**).**

The process of synthesis of proteins involves one of the central dogma of molecular biology, according to which genetic information flows from nucleic acids to proteins. It was first proposed by Crick in the year 1958. The first step of this central dogma is the synthesis of RNA from DNA. This is known as transcription. The second step involves a change of code from nucleotide sequences to amino acid sequences and is called translation.

It can be illustrated as follows:

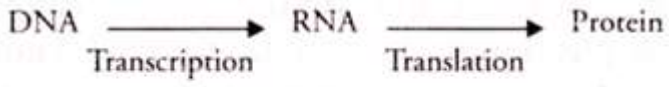


Fig. 9 Central dogma of molecular biology.

The DNA found in organisms has two main functions – replication and phenogenesis. Phenogenesis is a mechanism by which the phenotype of an organism is produced under the control of DNA in a given environment. The environment includes external factors such as temperature, quality and quantity of light, and internal factors such as hormones and enzymes.

The phenotype of an organism is the result of various embryological and biochemical activities of its cells from the zygotic to the adult stage. All these activities involve the action of a variety of structural and functional enzymes. The enzymes perform catalytic functions causing the splitting or union of various cellular molecules. Each reaction occurs in a stepwise manner involving the conversion of one substance to another.

The various steps involve the transformation of a precursor substance to its end product which ultimately is a structural or functional phenotypic trait. The various steps constitute a biosynthetic pathway. Each step of the pathway is catalysed by a specific enzyme, which in turn is produced by a specific gene.

DNA however, is not involved directly in the biosynthetic pathway. An intermediate molecule called mRNA is involved in the assemblage of amino acids to form enzymes. Thus, to produce a particular phenotypic trait, DNA transcribes mRNA which translates into either an enzymatic or structural protein. The groundwork for a functional relationship between genes and enzymes was laid in 1902 when Bateson reported a rare human defect known as alkaptonuria, which is inherited as a recessive trait.

The ***adaptor hypothesis*** is part of a scheme to explain how information encoded in DNA is used to specify the amino acid sequence of proteins. It was formulated by Francis Crick in the mid-1950s, together with the central dogma of molecular biology and the sequence hypothesis. It first appeared in an informal publication of the RNA Tie Club in 1955 and was formally published in an article “On Protein Synthesis” in 1958.

Explanation

The adaptor hypothesis was framed to explain how information could be extracted from a nucleic acid and used to put together a string of amino acids in a specific sequence, that sequence being determined by the nucleotide sequence of the nucleic acid (DNA or RNA) template. Crick proposed that each amino acid is first attached to its own specific “adaptor” piece of nucleic acid (in an enzyme-catalysed reaction). The order of assembly of the amino acids is then determined by a specific recognition between the adaptor and the nucleic acid which is serving as the informational template. In this way the amino acids could be lined up by the template in a specific order. Coupling between adjacent amino acids would then lead to the synthesis of a polypeptide whose sequence is determined by the template nucleic acid.

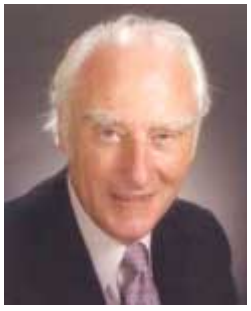
Basis

Crick’s thinking behind this proposal was based on a general consideration of the chemical properties of the two classes of molecule — nucleic acids and proteins. The amino acids are characterised by having a variety of side chains which vary from being hydrophilic to hydrophobic: their individual characters reside in the very different properties these side chains have. By contrast, a nucleic acid is composed of a string of nucleotides whose sequence presents a geometrically defined surface for hydrogen bonding. This makes nucleic acids good at recognising each other, but poor at distinguishing the varied side chains of amino acids. It was this apparent lack of any possibility of specific recognition of amino acid side chains by a nucleotide sequence which led Crick to conclude that amino acids would first become attached to a small nucleic acid — the adaptor — and that this, by base-pairing with the template (presumably as occurs between DNA strands in the double helix), would carry the amino acids to be lined up on the template.

Proof

That such adaptors do exist was discovered by Mahlon Hoagland and Paul Zamecnik in 1958. These “soluble RNAs” are now called transfer RNAs and mediate the translation of messenger RNAs on ribosomes according to the rules contained in the genetic code. Crick imagined that his adaptors would be small, perhaps 5-10 nucleotides long. In fact, they are much larger, having a more complex role to play in protein synthesis, and are closer to 100 nucleotides in length.

The discovery of messenger RNA (mRNA) by Sydney Brenner (1927-), Francis Crick (1916-), Francois Jacob (1920-) and Jacques Monod (1910-1976).
Once it became clear that genes are activated to make useful proteins, it became of the greatest interest to discover the molecular machinery involved. But evidence was confusing, and by the late 1950s the question of how exactly information from DNA was translated into proteins loomed as a roadblock for molecular biology.



Francis H. C. Crick
Courtesy The Salk Institute

The problem of the "missing messenger" was solved with a combination of experiment and collective insight about the role of ribonucleic acid (RNA). The close chemical kin to DNA—the principal difference is that uracil, rather than thymine, is one of the bases—RNA was known to play at least one role in protein synthesis. RNA-containing molecules, known as ribosomes, were found in the cytoplasm of cells, and protein synthesis could not proceed without them. But it remained unclear how ribosomal RNA received specific information from DNA.

Francis Jacob

In this regard, experiments with *E. coli* bacteria, conducted at the Institut Pasteur, became the focus of intense interest in 1959. The "PaJaMo" experiments—performed by Arthur *Pa*rdee, Francis *Ja*cob, and Jacques *Mo*nod—built upon research into the system of bacterial enzyme production pioneered by Jacques Monod. They involved observations of carefully controlled gene transfer during conjugation—mating between "male" and "female" bacteria.

In previous experiments, Monod had learned how to genetically manipulate the compounds that control sugar metabolism in *E. coli*—collectively known as the B-galactosidase system. He had first bred mutated "female" bacteria in which this system ceased to function. When normal "male" bacteria then penetrated and inserted genes into such bacteria, however, the system was immediately—within minutes—restored to normal and the bacteria could digest sugar. How such information transfer could take place so quickly suggested the existence of a specific, relatively simple molecule that was complementary to DNA.

Discussions among Monod, Jacob, Crick, and Brenner led to a solution. They recalled research from the early 1950s with bacteriophages—viral parasites that invade bacteria. Experiments had shown that soon after bacteriophages insert their DNA into bacterial cells, traces of RNA rapidly appear. In addition, the composition of such RNA closely resembled the DNA of the invading bacteriophage.

Jacques Monod
Courtesy the Archives, California Institute of Technology

With this as context, the PaJaMo experiments suggested that another type of RNA was rapidly synthesized from DNA. Comparatively short-lived, its crucial presence had been initially overlooked. But in 1960, Francis Jacob and Jacques Monod named this hypothetical molecule "messenger RNA" (mRNA). Its presence was subsequently confirmed by experiment.

As it was finally understood, several types of RNA represent a basic division of labor in protein synthesis. Messenger RNA (mRNA) presents information contained in DNA sequences to the ribosomes, which are structured by ribosomal RNA (rRNA). Other molecules, known as transfer RNA (tRNA), attach to specific amino acids and conduct them to the ribosomes for protein synthesis.

Properties of Genetic Codons:
As the genetic codon is read on mRNA, it is described in terms of four kinds of bases A, G, C, U. The sequence of template strand of DNA is read in the direction of 5′ → 3′.

1. Genetic codon is a triplet codon:
The consecutive three nucleotides of the coding strand of DNA code for one amino acid.

2. Redundancy of the code:
Out of 64 codons, 61 codons represent amino acids, the remaining three are stop codons. As there are only 20 amino acid, coded by 61 codons, several codons specify the same amino acids. In this way the codons are synonyms.

This phenomenon is called redundancy of the code or degenerate code. Except for methionine and tryptophan all other codons are multiple codons. Each of the three amino acids — leucine, serine and arginine is represented by six different codons.

Number of codons coding for different amino acids:

	Amino acids	Number of codons
1.	Methionine, Tryptophan	1
2.	Histidine, Glutamine, Lysine, Aspartic acid, Phenylalanine, tyrosine, asparagine, glutamic acid, cystine.	2
3.	Isoleucine	3
4.	Proline, Valine, Theonine, glycine, Alanine	4
5.	Leucine, serine, arginine	6

Wobble Hypothesis:

This was put forward by Francis Crick in 1965. According to this, hydrogen bonding between the codon of mRNA and anticodon of tRNA, there is a strict base pairing rule only for first two bases of the codon, while the base pairing involving the third base of codon appears to be less important. This is known as wobble hypothesis.

The first two bases of each codon are primary determinants of specificity. The third base pairing is not very stable and wobbles. For example, CUU, CUG, CUC, CUA codons, which differ only at the third base represent the same amino acid leucine. The first two bases of the codon form strong base pairs with the corresponding bases of the anticodon but the third base forms weak hydrogen bond.

At third position even unusual base pairing which does not conform to Watson and Crick base pairing rule can occur. Like adenine, cytosine and Uralic from A-I. C-I and U-I base pairs respectively at the third position, where I is inosine base.

Several codons meant for the same amino acids are recognized by the same tRNA. In this way a minimum of 32 tRNAs are required to translate 61 codons.

3. The codons are non-overlapping:

The same base cannot be a part of the two consecutive codons. They lie adjacent to each other.

4. The codons are comma less:

The three bases on DNA code for one amino acid and next three bases will code for next amino acid and so on. There is no gap or pause between the consecutive triplets.

5. Start codons:

The codons which initiate protein synthesis is called the start codon. The first amino acid of the polypeptide chain is always methionine coded by AUG codon. Therefore AUG is the start codon. Rarely the first amino acid is valine coded by GUG. In prokaryotes AUG codon codes for a modified amino acid, formyl methionine (f-Met).

6. Stop codons:

Protein synthesis stops before UAA codon, UAG codon and UGA codon. This indicates that these three codons are stop codons. They terminate the protein synthesis. The completed

polypeptide chain is released. Release factors (RF) enter the 'A' site of the ribosome and trigger hydrolysis of the peptidyl-tRNA occupying the 'P' site resulting in the release of newly synthesized protein.

Stop codons are also called non-sense codons. No tRNA can bind to these codons. UAG is called amber codon, UAA is called ochre codon and UGA is called opal codon.

7. Genetic code is universal:
A particular codon codes for the same amino acid in all organisms from prokaryotes to plants and animals including viruses.

The universality of genetic code provides strong evidence that life on the earth started from a common ancestor. When living forms appeared on the earth, the genetic code was established. It has not changed since then, throughout the evolution of living forms and has been preserved throughout the biological evolution.

8. Co-linearity:
The sequence of codons on mRNA and the sequence of corresponding amino acids in polypeptide chain are co-linear.

9. Translation of mRNA occurs in 5′ → 3′ direction.
Codons on mRNA and anticodons n tRNA are written as follows:

```
    1 2 3
5′ A C G 3′  codon on mRNA

3′ U G C 5′  Anticodon on tRNA
  3 2 1
```

First base of the codon pairs with the third base of the anticodon. Codons are written in 5′ 3′ but anticodon sequence is written with a backward arrow.

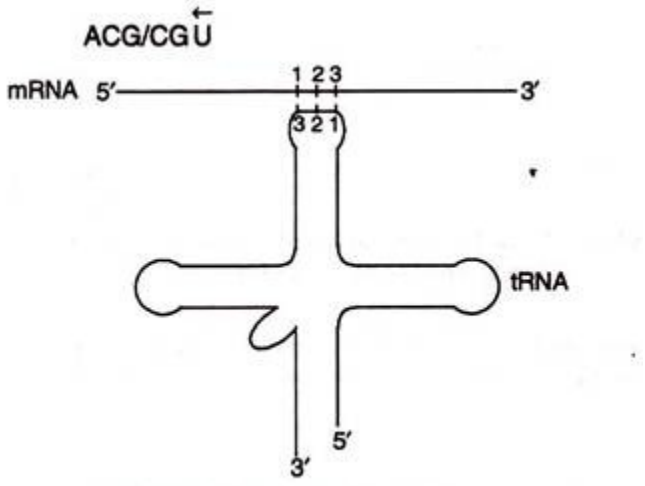


Fig. 11.2. Codon-anticodon base pairing

Effects of Mutation on Genetic Code:
A mutation causes changes in DNA sequence and this change is reflected in the sequence of RNA and then corresponding protein.

The main two types of mutation are:

1. Point Mutations:

These involve a single nucleotide.

2. Mutations Involving Longer Segments of Gene:

These include deletions of whole genes, translocations, transposable elements etc.

There are following four types of point mutation:

1. Silent Mutations:

Here, there is change of nucleotides but not of amino acids, because they affect the third base of the codon which is usually less important in coding. For example if CCA is changed to CCU, the amino acid coded will still remain the same i.e., proline.

2. Mis-sense Mutations:

These mutations change the meaning of codon, substituting one amino acid by another amino acid. For example, in human genetic disease sickle cell anemia, glutamic acid in the p-globin subunit of hemoglobin is replaced by valine. This changes the shape of the RBC.

3. Non-sense Mutations:

These arise when a codon for a codon is mutated into the termination codon which are UAG, UAA or UGA resulting in the termination of the polypeptide chain. This leads to the production of shorter (truncated) protein. For example, if U is replaced by G at position number 12th it will give

AUG	UAU	CCA	UA**G**	CCA	UAG
f-Met	Tyr	Pro	Stop codon		

4. Frame-shift Mutation:

These arise from the insertions or deletions of individual nucleotides and cause the rest of the message downstream the mutation to be read differently, producing an incorrect protein from that point onwards.

Wild type	GUG	GUG	GUG	GUG	GUG	GUG	
	Val	Val	Val	Val	Val	Val	
	Insertion of C ↓						
Insertion to the wild type	GUG	GUG	GCU	GGU	GGU	GGU	G
	Val	Val	Ala	Gly	Gly	Gly	
	Deletion of G ↑						
Deletion from the wild type	GUG	GUG	GUG	UGG	UGG	UG	
	Val	Val	Val	Tryp	Tryp		

5. Transition:

It involves substitution of one purine by another purine and substitution of one pyrmidine by another pyrimidine, for example, cytosine changes to uracil by oxidative deamination.

6. Transvertion:
In this case pyrimidine is replaced by a purine or a purine is replaced by a pyrimidine. Here A-T pair becomes T-A or C-G pair.

Triplete binding assay:

The **Nirenberg and Leder experiment** was a scientific experiment performed in 1964 by Marshall W. Nirenberg and Philip Leder. The experiment elucidated the triplet nature of the genetic code and allowed the remaining ambiguous codons in the genetic code to be deciphered.

In this experiment, using a ribosome binding assay called the triplet binding assay, various combinations of mRNA were passed through a filter which contained ribosomes. Unique triplets promoted the binding of specific tRNAs to the ribosome. By associating the tRNA with its specific amino acid, it was possible to determine the triplet mRNA sequence that coded for each amino acid.

Oswald Avery discovered that the substance responsible for producing inheritable change in the disease-causing bacteria was neither a protein nor a lipid, rather deoxyribonucleic acid (DNA). He and his colleagues Colin MacLeod and Maclyn McCarty suggested that DNA was responsible for transferring genetic information. Later, Erwin Chargaff discovered that the makeup of DNA differs from one species to another. These experiments helped pave the way for the discovery of the structure of DNA. In 1953, with the help of Maurice Wilkins and Rosalind Franklin's X-ray crystallography, James Watson and Francis Crick proposed DNA is structured as a double helix.[1]
In the 1960s, one main DNA mystery scientists needed to figure out was in translation how many bases would be in each code word, or codon. Scientists knew there were a total of four bases (guanine, cytosine, adenine, and thymine). They also knew that were 20 known amino acids. George Gamow suggested that the genetic code was made of three nucleotides per amino acid. He reasoned that because there are 20 amino acids and only four bases, the coding units could not be single (4 combinations) or pairs (only 16 combinations). Rather, he thought triplets (64 possible combinations) were the coding unit of the genetic code. However, he proposed that the triplets were overlapping and non-degenerate.[2]
Seymour Benzer in the late 1950s had developed an assay using phage mutations which provided the first detailed linearly structured map of a genetic region. Crick felt he could use mutagenesis and genetic recombination phage to further delineate the nature of the genetic code.[3] In the Crick, Brenner et al. experiment, using these phages, the triplet nature of the genetic code was confirmed. They used frameshift mutations and a process called reversions, to add and delete various numbers of nucleotides.[4] When a nucleotide triplet was added to or deleted from the DNA sequence, the encoded protein was minimally affected. Thus, they concluded that the genetic code is a triplet code because it did not cause a frameshift in the reading frame.[5] They correctly concluded that the code is degenerate, that triplets are not overlapping, and that each nucleotide sequence is read from a specific starting point.

Nirenberg and Leder could not decode the remaining codons in the same manner as Nirenberg did with Matthaei. Because the mRNA bases were taken up at random by the ribosome, it is hard to determine which specific codon correlates with the amino acid. For example, to pick out the correct codon among non-repeating codons (UCU, CUU, UUC) was difficult because they couldn't determine the specific sequence.Instead, Leder and Nirenberg used very short artificial RNA sequences (three nucleotides) in the cell-free systems. These shorter length fragments were long enough to allow the ribosome to bind with the type of tRNA molecule that is complementary to the one codon and still be detectable. The key step of the experiment was that

they labeled one type of amino acid at a time and then put the mixture through a Millipore filter (a type of nucleopore filter).This special filter allowed unbound tRNAs to pass through but did not allow the ribosomes with the bound triplet to pass through.The sample was then tested for radioactivity. If there was radioactivity found in the sample that did not pass through the filter, the corresponding amino acid had been added.

By the Cold Spring Harbor Symposium of 1966, between Nirenberg and Khorana the genetic code was almost completely decoded. Nirenberg was awarded the 1968 Nobel Prize in Physiology or Medicine. He shared the award with Har Gobind Khorana of the University of Wisconsin and Robert W. Holley of the Salk Institute. Working independently, Khorana had mastered the synthesis of nucleic acids, and Holley had discovered the exact chemical structure of transfer-RNA.

The New York Times reported on Leder and Nirenberg's discovery by explaining that "the science of biology has reached a new frontier," leading to "a revolution far greater in its potential significance than the atomic or hydrogen bomb."[2] Most of the Scientific community saw these experiments as highly important and beneficial. However, there were some who were concerned with the new era of Molecular Genetics. For example, Arne Wilhelm Kaurin Tiselius, the 1948 Nobel Laureate in Chemistry, asserted that knowledge of the genetic code could "lead to methods of tampering with life, of creating new diseases, of controlling minds, of influencing heredity, even perhaps in certain desired directions."[9]