

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

HANDOUT

LECTURE-02

CENTRAL DOGMA: BASICS OF DNA, RNA AND PROTEIN

Slide 1

Today we will talk about central dogma: basics of DNA, RNA and proteins.

Slide 2

Lecture outline

- the central dogma
- the structure and function of the DNA molecule
- structure and function of different types of RNA molecules
- proteins

Slide 3

Central Dogma

- Let's talk about central dogma which is information flow from DNA to RNA to protein. Let's say we want to make a building in IIT campus which is in the Powai area of Mumbai.
- The left side shows the map which shows that where this area is. DNA has a similar function, it is the genetic blueprint which contains only the information. Now, once the site is decided then a map has to be created for the building.
- RNA is molecular photocopy which can be used on the site.
- Now the building has to be prepared, to do that something similar in the body is proteins which are the building material. Now you need mortar and different types of brick to make the building and similar type of function being performed by the proteins in body.
- If you look at it this way, DNA is providing the genetic blueprint, RNA is providing the molecular photocopy and proteins are providing the building material.

Slide 4

- This orderly and unidirectional flow of information which is encoded in the base sequence of cells from DNA to RNA to protein that is known as central dogma.

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

- Although there are many pieces of evidence that challenges the linear logic of central dogma. But based on premise that DNA encodes mRNA and mRNA encodes protein, one cannot deny the conclusion that genes are the blueprint for the life and proteins are effector molecules.
- Due to this fact, the central dogma has guided research at the systems level.

Slide 5

Before we move on to discussion about structure and function of DNA molecules. Let me take you to some of the historical perspectives: the milestone discoveries which are related to DNA. **Slide 6**

Historical Milestones in the study of the DNA molecule

- First one is Mendel's law of genetics-proposed in 1865. Mendel proposed certain fundamental laws of genetics. The discrete factors which are now known as genes can transmit characteristics from one generation to next generation. a diploid individual must contain two copies of gene. Each parent can transmit one copy to the next generation.

Slide 7

- After this hereditary law from Mendel, lot of research started in this area and then one of the major milestones was DNA double helical structure which was discovered Watson and Crick in 1953. Watson and Crick published a paper in Nature in 1953 and they described "we wish to suggest structure for the salt of de-oxy-ribonucleotide: DNA. This structure has novel features which are of considerable biological interest". From that time the structure of DNA and function of DNA has been a subject of great research interest in the field of biology.

Slide 8

- In 1966, Nirenberg, Khorana and Holly determined the genetic code.

Slide 9

- Another major milestone discovery was recombinant DNA technology this was developed in 1972 by Cohen and Boyer.

Slide 10

NPTTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

- In 1977, the DNA sequencing methods were provided by Sanger, Maxam and Gilbert.

Slide 11

- Now let's move to the 1990s. One of the major interesting areas of research in biology was cloning. cloning is producing a cell or organism with the same nuclear material as another cell or organism. Dr. Ian Wilmut of Roslin Institute and his colleague cloned a sheep (Dolly) which was first large animal to be cloned from somatic cells.

Slide 12

- During this time the Human Genome Project was initiated and finally was completed in 2003. This was an initiative from international human sequencing consortium as well as Celera genomics.

Slide 13

- And now , very recently, different types of next generation sequencing approaches have made sequencing fast and affordable that it is revolutionizing the biological research at the genomic level.

Slide 14

I have given you a perspective of the historical background. Let's learn some of basics of DNA, RNA and Proteins.

- Deoxyribonucleic acid (DNA) which stores and transfers all the genetic information is a long polymer of nucleotide monomers that assumes a complex double helical structure.

Slide 15

- Main function is storage & transmission of genetic information, and there are two classes distinguished based on type of carbohydrate they contain - Deoxyribonucleic Acid (DNA) & Ribonucleic Acid (RNA).
- Nucleoside is when sugar and base is linked together. When phosphate binds to nucleoside, it gives rise to nucleotide. Then, these nucleotides are joined by phosphodiester bond to each other; it gives rise to nucleic acids.

Slide 16

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

The three dimensional structure illustrates close connection between molecular form and function

what are the basic components of DNA? We just talked about nucleotides which are a subunit of nucleic acids. A nucleotide consists of nitrogen containing base and five carbon sugar and phosphate groups. The sugar and phosphate molecules play a crucial role in forming linear sequence or structure.

Slide 17

Structure of nucleotides: the basic of building blocks of DNA

- DNA is made up of three basic components - a sugar, a nitrogenous base and a phosphate group.
- The sugar and base are linked to form a nucleoside and attachment of the phosphate group results in a nucleotide.
- Many such nucleotide units are linked together by means of a covalent bond known as the phosphodiester bond. This is formed between the 3' carbon of one sugar and 5' carbon of the next sugar via a phosphate group to give rise to a polynucleotide chain.
- DNA is composed of four different nitrogenous bases that are derivatives of the heterocyclic, aromatic compounds, purines and pyrimidines.
- Adenine and guanine are purines while thymine and cytosine are the pyrimidines. The nucleosides of these bases are known as deoxyadenosine, deoxyguanosine, thymidine and deoxycytidine respectively.

Slide 18

- Let's now talk about DNA double helix structure. As briefly discussed Watson and Crick in 1953, deduced the arrangement of two strands of DNA and proposed three dimensional structures. The double helix structure is composed of two DNA strands which are anti-parallel and non-covalently attached to each other.
- The sugar and phosphate backbone lies on outside, as you can see. A is going to pair with T and G is going to pair with C which is very specific base pairing. It is shown by red dashed line for the hydrogen bond.

Slide 19

DNA double helix structure-

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

- Hydrogen bonding between the complementary bases of the two strands of DNA holds them together, with A and T being held together by 2 hydrogen bonds and G and C by 3 bonds.
- This base pairing is often referred to as Watson-Crick pairs, named after the molecular biologists instrumental in elucidation of the structure of the double helix.
- The strands are oriented anti-parallel to each other and twist around an imaginary axis to form the double helical structure.
- The process by which two DNA strands of a double helix separate from one another by means of breaking of hydrogen bonds is known as DNA melting or denaturation.
- Heating of DNA solution causes the strands to separate and the temperature at which half of the DNA strands are in the double helical state while the remaining half are in random coil configuration is known as the melting temperature. The length of the nucleotide sequence & composition of DNA determines the T_m .
- The two sugar phosphate backbones of the DNA double helix are not equally spaced along the helical axis. This results in formation of grooves of unequal sizes between the backbone.
- The wider of the two grooves is known as the major groove while the narrower one is called the minor groove.
- Chargaff's rule states that DNA from any organism must have a 1:1 ratio of purine and pyrimidine bases.
- More specifically, it states that the amount of adenine is always equal to the amount of thymine and amount of guanine is equal to cytosine.

Three forms of DNA

- DNA exists in many possible conformations that include A-DNA, B-DNA, and Z-DNA forms. A and B forms are right handed helices whereas Z-DNA is a left handed helix. There are 10.9, 10.0 and 12.0 base pairs per helix turn in A, B and Z-DNA forms respectively. They differ in their overall structural proportions as well in the proportions of their major and minor grooves.

Slide 20

DNA function

- DNA carries the blueprint of life .

NPTTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

- In addition to transmitting hereditary information from one generation to the next by means of replication, the genes of DNA stores the code for protein sequences in all organisms.

Slide 21

- Replication: A fundamental process that occurs in all living organisms to transmit their genetic material from one generation to the next.
- Two copies of nucleic acid are synthesized from one parent molecule during the process of cell division such that each daughter cell obtains one copy of the genetic material.
- Base-pairing is a very important feature of DNA-self complementarity.
- During DNA replication, a DNA molecule is separated into two strands-each strand can act as template for generation of its partner strand.

Slide 22

Proposed models for DNA replication

- Conservative model - According to the conservative model, the two parental strands of DNA as a whole serve as a template for the synthesis of progeny DNA molecules. Thus, one of the daughter DNA molecules is actually the parental DNA while the other daughter DNA consists of two newly synthesized strands from fresh nucleotides.
- The dispersive model of DNA replication hypothesizes that the parental DNA molecule is cleaved into smaller double stranded DNA segments which serves as the template for synthesis of new DNA strands. The segments then reassemble into complete DNA double helices, with parental and daughter DNA segments interspersed. The content of parental DNA in the double helix goes on decreasing with each generation.
- According to the semi-conservative model of replication, each parental strand acts as a template for the synthesis of a new strand of DNA which is complementary to the parental strand. Each daughter DNA molecule always has one parental DNA strand and one newly synthesized daughter strand.
- Of the three replication models suggested, Meselson and Stahl proved that the semiconservative model was correct. For this they grew *E. coli* cultures for several generations in ^{15}N -containing medium that the bases in DNA contained ^{15}N instead of ^{14}N . Next they transferred & grew the cultures for several generations in a ^{14}N -containing medium. Throughout the period of growth,

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

samples were taken, cells lysed and the DNA analyzed by centrifugation in CsCl gradient. The parent DNA showed 1 band in CsCl gradient corresponding to ^{15}N DNA, the 1st generation daughter molecules also showed 1 band which was not at the same position as parent DNA.

- This corresponded to ^{14}N - ^{15}N DNA while the 2nd generation showed 2 bands, one of ^{14}N - ^{15}N and the other of ^{14}N light DNA. These results exactly matched the semiconservative replication model.

Slide 23

To understand about DNA replication, I think it's important to know few terms like-

Template: A polynucleotide DNA strand that serves as the guide for making a complementary polynucleotide.

Origin of replication: Unique sequences in the genome where replication is initiated.

Replication fork: The point where the two parental DNA strands separate to allow replication.

Helicase: An enzyme that unwinds a polynucleotide double helix using energy derived from ATP hydrolysis.

Primase: catalyzes synthesis of small pieces of RNA complementary to single stranded DNA that provides the free 3' OH end needed for DNA replication to begin.

DNA polymerase: synthesizes DNA by linking together deoxyribonucleoside monophosphates in an order directed by the complementary sequences of nucleotides in a template strand. DNA polymerases add nucleotides only on to a pre-existing 3'-OH group. Prokaryotes have three types of DNA Polymerase while eukaryotes have five.

Leading and lagging strands: DNA polymerase can synthesize DNA only in 5'to 3' direction. Therefore, it synthesizes one strand (leading strand) continuously and the other strand(lagging strand) discontinuously. Each new piece synthesized on the lagging strand template is called Okazaki fragment.

DNA ligase: catalyzes formation of a phosphodiester bond between 5' phosphate of one strand of DNA and 3' hydroxyl of another thereby covalently linking DNA fragments together during DNA replication and repair.

Replication process: DNA double helix unwinds at replication fork. Two single strands are produced which serve as templates for polymerization of free nucleotides. DNA

NPTTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

polymerase polymerizes these nucleotides by adding new nucleotides to the 3' end of DNA chain. Since it adds only at 3' ends, polymerization on one template is continuous, and produces the leading strand. On the other, it is in short stretches and discontinuous on lagging strand, known as Okazaki fragments. Okazaki fragment primed by a short RNA primer (synthesized by primase), provides a 3' end for deoxyribonucleotide addition.

Slide 24

- Where and when replication takes place and multiple events involved in this process have to be accurate for replication to occur. Let me demonstrate how DNA replication works in following animation

DNA replication-

- DNA undergoes semi-conservative, bi-directional replication which begins with the unwinding of the DNA double helix.
- This is done by the enzyme DNA helicase which binds to the replication fork and unwinds the DNA using the energy of ATP hydrolysis.
- As this occurs, the enzyme DNA gyrase relieves the torsional strain that builds up during the process in the unwound part of the double helix.
- The single-stranded binding proteins bind to and stabilize the unwound single stranded regions of the DNA helix to allow replication to occur.
- Initiation of DNA replication is carried out by a primase enzyme which synthesizes short RNA primer fragments since DNA Polymerase is not capable of carrying out this process.
- The SSBs are displaced as the short fragments get synthesized.
- Synthesis takes place in the 5' to 3' direction such that nucleotides can be added to the free 3' OH group with concomitant cleavage of the high energy phosphate bond of the incoming nucleotide.
- Elongation takes place continuously in the 5'-3' direction on one strand, known as the leading strand.
- On the other strand, replication is discontinuous with short primers being added as the helicase unwinds the double helix. Elongation is carried out by DNA Pol III, a highly processive enzyme.
- The short fragments synthesized on the lagging strand are known as Okazaki fragments. The entire DNA is unwound in this manner by DNA helicase with DNA Pol III synthesizing the new complementary strands.
- The RNA primers are then removed and the gaps filled by the enzyme DNA Pol I. The Okazaki fragments on the lagging strand, which still have a nick between

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

two consecutive fragments, are then joined together by means of the enzyme DNA ligase.

- Sealing of the nicks completes the process of replication after which all the machinery dissociates from the DNA strands.

Slide 25

DNA Transcription

- Transcription is a process by which information from a double stranded DNA molecule is converted to a single stranded RNA molecule by making use of one strand as the template.
- The process differs slightly between prokaryotes and eukaryotes. First talk about prokaryotic transcription.
- In prokaryotes, RNA polymerase initiates transcription by binding to DNA at promoters which contain specific sequences at -35 and -10 bases before the transcription start site at +1. RNA polymerase locally unwinds the DNA after binding and starts incorporation of ribonucleotides which are complementary to the template DNA strand.
- The chain grows in 5-to-3 direction until intrinsic or rho dependent mechanism dissociates polymerase and RNA from template DNA.

Slide 26

- The eukaryotic transcription involves initiation, elongation, and termination phase of RNA synthesis. It has similarity with prokaryotes but there are few differences.
- 3 types of RNA polymerases; but only RNA polymerase II transcribes mRNAs and coordinates the numerous events of RNA synthesis and processing.
- RNA polymerase II does not bind directly to promoter DNA, but rather to general transcription factors, one of which recognizes TATA sequence in eukaryotic Promoter
- Let's look at the prokaryotic and eukaryotic transcription in more detail in animation

Slide 27

Animation

Transcription of DNA

- Transcription is the process by which information from a double stranded DNA molecule is converted to a single stranded RNA molecule.

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

- For prokaryotic transcription to begin, the RNA polymerase holoenzyme consisting of the core enzyme bound to the σ factor must bind to the promoter region.
- The σ factor is responsible for recognition of the promoter sequence. Binding results in a local unwinding of around 17 base pairs centered on the promoter.
- At this point RNA polymerase is correctly oriented to begin transcription from the +1 nucleotide.
- In case of eukaryotic transcription initiation, RNA polymerase binds to the promoter region along with several transcription factors (TF), which recognize the promoter site.
- The first step is the binding of TFIID. This complex acts as a binding site for TFIIB which then recruits RNA polymerase II and TFIIH. Finally TFIIE and TFIIH also bind to produce the complete transcription initiation complex. In prokaryotes, the sigma factor dissociates from the core enzyme, a process known as promoter clearance, once it has synthesized around 9-10 nucleotides.
- RNA polymerase continues elongation of new RNA chain in the 5'-3' direction by unwinding the DNA ahead of it as it moves and re-winding the DNA helix that has already been transcribed.
- Eukaryotic elongation is similar except that the polymerase involved is RNA polymerase II. Termination of transcription is signaled by controlling elements called terminators that have specific distinguishing features.
- Prokaryotic termination can be rho-dependent or rho-independent. In rho-dependent termination, one subunit of the rho protein gets activated by binding to ATP after which the other subunit binds to the RNA transcript and moves to the stalled transcription complex.
- Hydrolysis of ATP leads to release of the RNA transcript as well as RNA Polymerase, thereby terminating the transcription process. Rho-independent termination takes place due to the formation of a hairpin loop structure by the newly synthesized RNA transcript.
- The terminators for this mechanism have two specific features – the first is a region on the template that will produce a self-complementary sequence on the RNA transcript located around 15--20 nucleotides before the expected end of the RNA.
- The next feature is a conserved sequence of 3 adenine residues on the template near the 3' end of the hairpin. Formation of the hairpin disrupts the weak AU interactions, thereby allowing dissociation of the newly synthesized RNA transcript and the RNA polymerase.

Slide 28

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

Let's talk about structure and function of RNA.

Slide 29

RNA: Structure and Function

- RNA is made-up of nucleotides: A,G,C and *Uracil (U)*.
- Difference from DNA is that Uracil replaces Thymine (T).
- RNA is synthesized using DNA molecule as template and RNA is an intermediate in flow of information from DNA to protein. A functional component of molecular machinery, called ribosome carries out the translation process.
- It has an important role in gene expression

Slide 30

RNA Structure and function-

Structure of RNA

- RNA is made up of three basic components - a sugar, a nitrogenous base and a phosphate group. The sugar and base are linked to form a nucleoside and attachment of the phosphate group results in a nucleotide. Many such nucleotide units are linked together by means of a covalent bond known as the phosphodiester bond.
- This is formed between the 3' carbon of one sugar and 5' carbon of the next sugar via a phosphate group to give rise to a polynucleotide chain. RNA is composed of four different nitrogenous bases that are derivatives of the heterocyclic, aromatic compounds, purines and pyrimidines.
- Adenine and guanine are purines while uracil and cytosine are the pyrimidines. RNA exists mainly as a single-stranded molecule.
- The base stacking interactions often tend to make the RNA assume a right-handed helical conformation. Single stranded RNA also forms secondary structures by folding back on itself resulting in formation of loops and hairpins due to base pairing interactions.
- Functional RNA molecules often require a specific tertiary structure, the scaffold of which is provided by the secondary structure. These RNA due to their large negative charge are stabilized by metal ions.

Slide 31

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

Slide 32

Different classes of RNA

- Messenger RNA is formed from a DNA template by transcription. This mRNA is often referred to as the pre-mRNA in eukaryotes since it undergoes further processing to form a mature mRNA. A fully processed eukaryotic mRNA includes a 5' cap, where the nucleotide at the 5' end is modified by addition of 7-methyl guanosine and a poly A tail at the 3' end which serves to protect the mRNA from degradation by exonucleases.
- The mRNA also contains 5' and 3' UTRs that contain signal sequences and serve as binding sites for various proteins. The coding sequence is flanked by start and stop codons that define the beginning and end of the gene to be transcribed.
- Longer RNA precursors are modified by enzymatic removal of nucleotides from the 5' and 3' ends to form the tRNA structure. Additional processing of the tRNA such as attachment of the 3' CCA trinucleotide unit and modification of certain bases takes place in certain bacteria and almost all eukaryotes. All tRNAs have a common secondary structure represented by a clover leaf having four base-paired stems. The anticodon loop recognizes the corresponding mRNA codon while the acceptor stem adds the suitable amino acid to the growing polypeptide chain.
- rRNA is the central component of the ribosome involved in protein synthesis in all living cells. Prokaryotic 70S ribosome is composed of 50S and 30S subunits where S is a measure of the rate of sedimentation of the respective components in a centrifuge. rRNAs are derived from longer precursors called pre-rRNA. A single 30S rRNA precursor is processed by several enzymes to give rise to 16S, 23S and 5S rRNAs in bacteria.
- Eukaryotic 80S ribosome is composed of 60S and 40S subunits where S is a measure of the rate of sedimentation of the respective components in a centrifuge. In eukaryotic vertebrates, a single 45S rRNA precursor is processed by several enzymes to give rise to 18S, 5.8S and 28S rRNAs.
- Different classes of mRNA: Basically three classes- mRNA, tRNA and rRNA
- Messenger RNA or mRNA- which is least the abundant around 5% of total RNA and it provides template for the protein synthesis or translation process.
- Transfer RNA or tRNA- it carries amino acids in an activated form to the ribosome which is in the intermediate abundance.
- Ribosomal RNA or rRNA- This is a major component of ribosome and fulfills catalytic and structural roles.

Slide 33

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

Function of different classes of RNA in protein synthesis

- mRNA: The messenger RNA is a long sequence of nucleotides that serves as a template for protein synthesis. It is transcribed from a DNA template by RNA Polymerase and gets translated into the amino acid sequence of the corresponding protein.
- Eukaryotic mRNA requires extensive processing to form the mature mRNA while prokaryotic mRNA does not require. Typical mRNA structure is composed of the following regions:
 - tRNA- A relatively small RNA molecule involved in protein synthesis that binds an amino acid at one end and base pairs with an mRNA codon at the other, thus serving as an adaptor that translates an mRNA code into a sequence of amino acids. A tRNA molecule consists of the following components:
 - rRNA- rRNA forms the central component of ribosomes. It has both catalytic and structural roles in protein synthesis. The ribosome that houses this rRNA consists of a large subunit and a small subunit.

Slide 34

Protein synthesis

- Initiation of protein synthesis is carried out by binding of the mRNA to the small ribosomal subunit such that its initiation codon, most often an AUG sequence, is aligned at the P site. The initiator tRNA that carried a modified methionine amino acid on its acceptor stem then binds to the ribosomal subunit by means of codon-anticodon interactions. The large subunit is then assembled on top of this to form the initiation complex. Other initiation factors are involved which ensure correct positioning of all the components.
- The next incoming aminoacyl tRNA carrying the amino acid corresponding to the next codon occupies the A site. A peptide bond is then formed between the amino acid in the A site and the P site with the P site amino acid being transferred to the A site. The unbound tRNA then leaves the P site and is moved to the exit or E site briefly before being removed.
- Once the peptide bond has been formed, the ribosome moves one codon towards the 3' end of the mRNA such that the tRNA in the A site now occupies the P site and the A site is again free for the next incoming aminoacyl tRNA. Multiple such rounds of elongation followed by translocation of the tRNAs are carried out to form the growing polypeptide chain.
- When the ribosome encounters the termination sequence, typically UAA, UAG, UGA, a release factor binds to the vacant A site and the polypeptide chain is hydrolyzed and released. Other termination factors aid this process. Once

NPTEL VIDEO COURSE – PROTEOMICS

PROF. SANJEEVA SRIVASTAVA

synthesis is complete, the ribosomal subunits dissociate from each other and all components are separated until commencement of the next round of translation.

Slide 35

in summary, today we talked about the importance of central dogma. We discussed the basics of DNA structure and function. We then talked about basics of RNA structure and function. We discussed transcription and translation process briefly. We did not have enough time to go through proteins but in the next lecture we will continue on amino acids and different level of protein structure in more detail.