Proteomics: Principles and Techniques Prof: Sanjeeva Srivastava Department of Biosciences and Bioengineering Indian Institute of Technology, Bombay

Lecture No. # 02 Central Dogma: Basics of DNA, RNA, Proteins

(Refer Slide Time: 00:29)



(Refer Slide Time: 00:38)



### (Refer Slide Time: 01:20)



Welcome to the proteomics course, today I talk about central dogma basics of D N A, R N A and proteins. So, lecture outline is we will briefly discuss, the central diagram the structure and function of D N A molecule and structure and function of different type of R N A and then we will talk about proteins. So, let us talk about central dogma, which is information flow from D N A to R N A to protein let us move on to make a building in I I T campus, which is in Powai area in Mumbai.

So, the left side shows the map, which shows that where this area is. So, D N A is doing something similar type of function D N A is the genetic blueprint, which contains only the information. Now, once the site is decided then map has to be created for the building. So, R N A is the molecular photocopy, which can be used on the site from the cell contractors. 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 water and different type of bricks to make the building and similar type of function being performed from the proteins in body. So, if you look at this way D N A is providing the genetic blueprint, R N A is providing the molecular photocopy and then proteins are providing the building material.

(Refer Slide Time: 02:22)



So, proteins can provide and transform the one dimensional information from the sequel to the three dimensional functional formation this orderly and unidirectional flow of information, which is encoded in the base sequence of cells from D N A to R N A and protein that is known as central dogma.

Although, now there are many evidences that challenges the linear logic of central dogma, but based on the previces that, D N A encodes m R N A, an m R N A encodes the protein one cannot deny the conclusion that genes are the blueprint for the life. And proteins are (( )) molecules due to this fact the central dogma has guided research at the systems level.

(Refer Slide Time: 03:25)



(Refer Slide Time: 03:42)



Before we move onto a discussing about the structure and function of D N A let me take, you to some of the historical prospective the milestone discoveries, which are related to the D N A. So, first one is Mendel law of genetics in 1865, Mendel gave very fundamental laws of genetics the discrete factors, which are now known as genes can transmit characteristics from one generation to the next generation.



So, deployed individual must contain two copies of genes each parent can transmit one copy to the next generation. After, this hereditary law from Mendel lot of research started in this area and then what are the major milestone was D N A double helix structure, which was discovered by Watson and Crick in 1953. Watson and Crick publish a paper in 1953 in nature and they described we are wish to suggest a structure for the results of de oxy ribonucleic acid D N A. This structure has novel features of which are of considerable biological interest from that time the structure and function of D N A has been a subject of great research interest in the field of biology. The shown in the picture is James Watson and I had an opportunity to meet with him in cold spring harbor.

(Refer Slide Time: 05:17)



(Refer Slide Time: 05:29)



(Refer Slide Time: 05:39)



So, I have shown that picture here by 1966 Nirenberg, Khorana and Holley they determined the genetic code another major milestone discovery was recombinant D N A technology, which was developing in 1972 by Cohen and Boyer. In 1976 the D N A sequencing method, where provided by Sanger, Maxan and Gilbert. Now, let move to the 1990s what are the major interesting area of research in biology was cloning.

(Refer Slide Time: 05:57)



(Refer Slide Time: 06:27)



(Refer Slide Time: 06:48)



### (Refer Slide Time: 07:08)



So, cloning is producing a cell or organism with the same nuclear material as another cell or organism. Doctor Ian Wilmut Roslin institute and his colleagues cloned a sheep dolly, which was first large cloned animal from the somatic cells. There in this time the human genome project was initiated and many of the genome projects and specially the human genome project was completed 2003, which was an initiative from international human genome sequencing consortium as well as, Celera genomics. And now, very recently different type of next generation sequencing approaches have made sequencing so much fast and affordable that it is revolutionizing the biological research at the genomic level. After giving you the prospective some of the historical background let us go to some of the basics of D N A, R N A and proteins.



Although this is not the directly linked to the proteomics, but to understand the concept at the systems level I think it is very important to refresh your fundamentals of D N A, R N A and proteins. de oxy ribonucleic acid or D N A, which stores and transmutes all the genetic information is a long polymer of nucleotide monomers that assumes a complex double helical structure. So, main function of nucleic acid is the storage and transmission of genetic information and there are two major classes D N A and R N A. The nucleoside is when you have a sugar and base, when phosphate molecule is also added that gives rise to nucleotide and, then nucleotides join by the phosphodiester bonds give rise to nucleic acid as you can see in this slide. (Refer Slide Time: 08:33)



(Refer Slide Time: 09:10)



So, what are the basic components of D N A? We just talked about nucleotide, which is a subunit of nucleic acid. So, it consist of nitrogen containing base 5 carbon sugar and phosphate groups. The sugar and phosphate molecules play crucial role in forming the linear D N A sequence or a structure. So, the three dimensional structure illustrate a very close connection between the molecular form and function of D N A. D N A 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 a means, of a covalent

bond known as phosphodiester bond. This is formed between the three prime carbon of one sugar and five prime carbon of the next sugar via a phosphate group to give rise to a polynucleotide chain.

(Refer Slide Time: 10:00)



(Refer Slide Time: 10:12)



(Refer Slide Time: 10:18)



(Refer Slide Time: 10:22)

| <b>~</b> | H H                               |
|----------|-----------------------------------|
| Cytosine | N<br>N<br>N<br>N<br>Deoxycytidine |
| NPTEL    |                                   |

### (Refer Slide Time: 10:50)



D N A is composed of four different nitrogenous bases, that are derivatives of the heterocyclic aromatic compounds purines and pyrimidines adenine and guanine are purines. The nucleosides of these bases are known as deoxyguanosine thymine and cytosine are pyrimidines the nucleoside of cytosine is known as deoxycytidine. Let us now talk about D N A double helix structure, as we briefly discussed Watson and Crick in 1953 deduced the arrangement of tools trends of D N A and proposed a three dimensional structure. The double helix structure is composed of two intervened strands, which are anti parallel and non covalently attached to each other.

(Refer Slide Time: 11:24)



# (Refer Slide Time: 11:40)



(Refer Slide Time: 12:05)



# (Refer Slide Time: 13:05)



# (Refer Slide Time: 13:12)



### (Refer Slide Time: 13:27)



## (Refer Slide Time: 14:10)



The sugar and phosphate backbone lies on the outside as you can see, the a is going to pair with t and g is going to pair with c, which is very specific base pairing and this are shown with the red dash line for the hydrogen bonds. Hydrogen bonding between the complimentary bases of two strands of D N A holds them together, with a and t being held together by two hydrogen bonds and g and c by three bonds. This base pairing is often referred to as Watson crick pair named after the molecular biologist who were is to mental in elucidation of structure of double helix.

The strands are oriented entire parallel to each other and twist around an imaginary axis to form the double helix structure. The process by which two D N A strands of a double helix separate from one another by means, of breaking of hydrogen bonds is known as D N A melting or denaturation. Heating of D N A solution causes, the strands to separate and the temperature at, which half of the D N A strands are in the double helical state. While the remaining half are in random coil configuration is known as, melting temperature.

The length of the nucleotide sequence and composition of D N A determines the t m the two sugar phosphate backbones of D N A, double helix are not equally spaced along the helical axis. This results in formation of grooves of unequal size between the backbone. The wider of the two grooves is known as, major groove, while the narrower one was called as minor groove. Chargaff's rule states, that D N A from any organism must have one to one ratio of purine and pyrimidine bases. More specifically it states that amount of adenine is always equal to the amount of thymine and amount of guanine is always equal to cytosine.

(Refer Slide Time: 14:15)



## (Refer Slide Time: 14:38)



(Refer Slide Time: 15:08)





Three forms of D N A, D N A exist in many possible conformation, that include A D N A, B D N A and Z D N A forms a and b are right handed helices, where as Z D N A is left hand helix. There are 10.9, 10 and 12 base pairs per helix turn in A B and Z D N A respectively. They differ in their over line structure prepositions as well as, in the proportions of their major and minor grooves. After knowing the basics of D N A structure, let us talk about function of D N A replication is fundamental process, that occurs in all living organisms to transmit the genetic material from one to the next generation.

The two copy of nucleic acid are synthesized from one parent molecule during this process of cell division in such a manner, that each daughter cell obtains one copy of genetic material. Now, when we looked at base pairing in d n a, that is one of the very important feature of D N A for the self complementarily. The D N A replication one D N A molecule will generated in to two strands and each strand can act as a template for generation of it is partner strand. The different types of models, which have been proposed for D N A replication including conservative model, dispersive model and semi conservative model.

## (Refer Slide Time: 16:56)



(Refer Slide Time: 17:40)



### (Refer Slide Time: 17:50)



(Refer Slide Time: 19:45)



According to the conservative model the two parental strands of D N A as a whole, serve as a template for the synthesis of pro gene D N A molecule. Thus, one of the daughter D N A molecule is actually the parental D N A, while other daughter D N A consist of two newly synthesized strands from fresh nucleotides. The dispersive model of D N A replication hypothesizes, that the parental D N A molecule is cleaved into smaller double standard D N A segments, which serve as the template for synthesis of new D N A strands. The segments then reassemble into complete D N A double helices with parental and daughter D N A segments inter (( )). The content of parental D N A in double helix decreases with each generation according to the semi-conservative model of replication. Each parental strand acts as a template for the synthesis of new strand of D N A, which is complementary to the parental strand. Each daughter D N A molecule always has one parental D N A strand and one newly synthesized daughter strand. Among the three replication models discussed, scientist Meselson and Stahl proved that the semi conservative model was most accurate for this they grew e coli cultured for several generations in fifteen n containing medium.

So, that the bases in D N A contained fifteen n instead of fourteen nitrogen. Next they transferred and grew the cultured for several generation in a fourteen nitrogen containing medium. Throughout the period of growth samples were taken cells sliced and D N A analyzed by centrifugation in caesium chloride gradient. The parent D N A showed one band in caesium chloride gradient corresponding to 15 N D N A. The first generation daughter molecule also showed one band, which was not at the same position as parent D N A. This corresponded to 14 N, 15 N hybrid D N A, while the second generation showed two bands one of 14 N, 15 N and other of 14 Nitrogen light D N A. These results exactly match the semi conservative replication model.



(Refer Slide Time: 20:18)

To understand about D N A replication I think it is important to know some of the terms, which are mentioned in the slide the template that is a polynucleotide D N A strand has guide for making a complementary polynucleotide. Origin of replication, that is the

unique sequence genome, where replication is initiated replication fork this is a point, where two parental D N A strands separate to allow the replication. Helicase this is enzyme, which helps in unwinding a polynucleotide double helix using the energy, which is derived from a t p hydrolysis primase.

This enzyme catalyses synthesis of small pieces of R N A, which is complementary to which single stranded D N A that provides the free three prime hydroxyl and required for the D N A replication to start noise. The leading and lagging strands the D N A polymerase can synthesize D N A only in five prime directions to three prime directions. Therefore, it is synthesizes one is strand, which is leading strand continuously and other strand, which is lagging strand discontinuously.

Each new piece is synthesize on the lagging strand template is known as, okazaki fragment D N A ligase enzyme. That catalyzes formation of phosphodiester bond between five prime phosphate of one strand of D N A and three prime hydroxyl of another thereby it covalent links the D N A fragment together during D N A replication process. Say summary the D N A replication process involves the D N A double helix that unwind at the replication fork. The two single strands are produced, which serve as template for the polymerization of free nucleotide.

The D N A polymerase, polymerases this nucleotides by adding new nucleotides to the three prime end of D N A chain. Since this acts only at the three prime end polymerization on one template discontinuous and produces leading strand on the other hand it is in the short stretches and discontinues on the lagging strand. Okazaki fragment prime by a short R N A primer, which is synthesize by enzyme primase, it provides a three prime end for the deoxyribonucleotide addition. So, where and when the replication takes place and multiple events, which are involved in this process have to be very accurate for replication to occur.

# (Refer Slide Time: 23:35)



(Refer Slide Time: 24:35)



# (Refer Slide Time: 25:43)



(Refer Slide Time: 25:45)



#### (Refer Slide Time: 26:08)



So, let us see how, D N A replication works in following animation D N A undergoes semi conservative by directional replication, which begins with unwinding of D N A double helix. This is done by enzyme D N A helicase, which binds to the replication fork and unwind the D N A using energy of a t p hydrolysis. As this occurs they enzyme D N A gyres release the torsional strain that builds up during this process in the unbound part of double helix. The single stranded binding proteins s s b's bind to and stabilize the unbound single stranded reasons of D N A helix to allow replication to occur. Initiation of D N A replication is carried out by a primes enzyme, which synthesizes short R N A primer fragments.

Since D N A, polymerase is not capable of carrying out this process the s s b's displaced as short fragments get synthesized. Synthesis takes place in five prime to three prime directions such that, the nucleotides can be added to free three prime hydroxyl groups. Elongation takes place continuously in five prime to three prime directions on one strand known as the leading strand. On other strand replication is discontinuous with short primers being added as the helicase unwinds the double helix. Elongation is carried out by D N A polymerase third a highly processivity enzyme, the short fragments synthesized on the lagging strand are known as okazaki fragments.

### (Refer Slide Time: 27:10)



The entire D N As unbound in this manner by D N A helicase with D N A polymerase third synthesizing the new complementary strands. The R N A primers are then removed and these gaps filled by enzyme D N A polymerase first. The okazaki fragments on the lagging strand, which is still have a nick between two consecutive fragments are then join together, by means of enzyme D N A ligase sealing of the nick complete the process of replication after which all the machinery disassociates from the D N A strands.

<section-header><figure><image>

(Refer Slide Time: 27:45)



Let us now, move on to transcription of D N A and first talk about prokaryotic transcription. Transcription is a process by, which information from a double stranded D N A molecule is converted to single stranded R N A molecule by making use of one strand as a template. This process differ slightly between prokaryote and eukaryotes more specifically in the prokaryotes it is done by the R N A polymerase, which initiates the transcription by binding at the promoter, which contain a specific sequences at minus thirty five and minus ten bases. Before the transcription is start at plus one R N A polymerase locally unwinds the D N A after binding and starts incorporation of ribonucleotides, which are complementary to the template D N A strand.

The chain grows in five to three prime direction until row dependent or independent mechanism disassociates the polymerase and R N A from the template D N A. The eukaryotic transcription it involves initiation elongation and termination phases of R N A synthesis, which are having very similarities with the prokaryotes. But which has few differences in the eukaryotic transcription there are three types of R N A polymerase, but only R N A polymerase to transcribes m R N A's and coordinates numerates events of R N A synthesis and processing. The only R N A polymerase to does not binding the directly to the promoted D N A, but rather to the transcription factors one of which recognizes the Tata sequence in eukaryotic promoter.

(Refer Slide Time: 30:06)



(Refer Slide Time: 30:34)

| Prokaryotic transcription initiation |  |
|--------------------------------------|--|
| 5'<br>3'<br>5'<br>3'                 |  |
| NPTEL                                |  |

# (Refer Slide Time: 31:08)



(Refer Slide Time: 31:38)

| Eukaryotic transcription initiation            |          |
|--|----------|
| Closed promoter complex<br>5' The Fib<br>3' 5' |          |
| NPTEL  | <u>.</u> |

### (Refer Slide Time: 31:43)



So, let us discuss the prokaryotic and eukaryotic transcription in more detail in following animation. Transcription is the process by which information form a double helix D N A molecule is converted to a single stranded R N A molecule for the prokaryotic transcription to begin the R N A polymerase hollow enzyme consisting of the core enzyme bound to the sigma factor must bind to the promoter region. This sigma factor is responsible for recognition of the promoter sequence. Thus, binding results local unwinding of around seventeen days periods centered around the promoter region.

At that point R N A polymerase is correctly oriented to begin transcription from the plus one nucleotide. In case, of eukaryotic transcription initiation the R N A polymerase to binds to the promoter region along with several transcription factors which recognizes the promoter tide. The first step is the binding of transcription factor second d this complex acts as a binding site for t f second b push them recruits R N A polymerase second and t f second f. Finally, the t f second e and t f second h also bind to complete transcription initiation process. (Refer Slide Time: 31:58)



(Refer Slide Time: 32:10)

| Elongation   |   |
|--|---|
| 5'<br>3'<br>Direction of<br>movement<br>DNA strand |   |
| NPTEL  | × |

(Refer Slide Time: 32:30)



(Refer Slide Time: 32:40)



In prokaryote the sigma factor dissociates from the core enzyme by a process known as, promoter clearance, once it has synthesized around nine to ten nucleotides. The R N A polymerase continues elongation of new R N A chain in five prime three plane direction by unwinding the D N A ahead of it as it moves and rewinding the D N A helix that has already been transcribed. Determination of transcription is signaled by controlling elements called terminators. That have a specific distinguishing features the prokaryotic termination can be rho dependent or rho independent.

# (Refer Slide Time: 33:18)



(Refer Slide Time: 33:34)



(Refer Slide Time: 33:37)



(Refer Slide Time: 34:29)



### (Refer Slide Time: 34:36)



In rho dependent termination one subunit of the rho protein gets activated by binding to ATP. After, which the other sub unit binds to the R N A transcript and move to the start transcription complex. The hydrolysis of a t p leads to release of the R N A transcript as well as, R N A polymerase thereby, terminating the transcription process. Another type of termination, which is rho independent termination, it takes place due to the formation of hairpin loop structure by newly synthesized R N A transcript. The terminators for this mechanism have two specific features. First is a region on template that will produce a self complimentary sequence on R N A transcript located around fifteen to twenty nucleotides before the expected end of R N A. The next feature is a conserved sequence of three adenine residual on template near the three prime end of hairpin.

## (Refer Slide Time: 34:53)



## (Refer Slide Time: 35:26)



### (Refer Slide Time: 36:03)



The formation of hairpin disrupts the weak a u interaction and allows dissociation of newly synthesized R N A transcript and R N A polymerase. Now, let us talk about a structure and function of R N A, R N A is made up of nucleotides a g c and uracil. So, difference from D N A s that uracil instead of thymine is present in R N A. Now, R N A is a synthesized using D N A s template molecule and R N A is intermediate in flow of information from D N A, which passes on to the proteins. The functional component of molecular machinery is known as ribosome's, which helps in the translation process R N A is made up of three basic components a sugar a nitrogenous base and a phosphate group.

# (Refer Slide Time: 36:30)



# (Refer Slide Time: 36:50)



### (Refer Slide Time: 36:56)



The sugar and base are length to form a nucleoside and attachment of the phosphate group results in a nucleotide. Many such nucleotide units are length together by means of a covalent bond known as phosphodiester bond. This is formed between the three prime carbon of one sugar and five prime carbon of next sugar via a phosphate group to give rise to a polynucleotide chain. R N A is composed of four different nitrogenous bases that are derivatives of heterocyclic, aromatic compounds, purines and pyrimidines. Purines include adenine and guanine.

(Refer Slide Time: 37:03)



(Refer Slide Time: 37:04)



(Refer Slide Time: 37:10)



### (Refer Slide Time: 37:25)



While, uracil and cytosine they are pyrimidines R N A exist mainly as a single stranded molecule. The base stacking interaction often tends to make R N A assume a right handed helical conformation. Single stranded R N A also forms secondary structure by folding back on itself resulting in formation of loops and hairpins due to base pairing interactions. The functional R N A molecule often require a specific tertiary structure the scaffold of, which is provided by the secondary structure. These, R N A due to their large negative charge are stabilize by metal ions.

(Refer Slide Time: 38:20)



### (Refer Slide Time: 40:00)



The messenger R N A is formed from a D N A template by transcription. This m R N A is often referred to as the pre m R N A in eukaryotes since it undergoes further processing to form a mature m R N A. A fully processed eukaryotic m R N A includes a five prime cap, where the nucleotide at the five prime end is modified by addition of seven methyl guanosine. And a poly a tale at the three prime end this use to protect the m R N A by degradation by exonucleases.

The m R N A also contains five prime and three primes U T R's that contain signal sequences and server binding site for various proteins. The coding sequence is flagged by start and a stop codons that define the beginning and end of the gene to be transcribed. The longer R N A precursor's r modified by enzymatic removal of nucleotides from five prime and three prime ends to form the t R N A structure. Additional processing of t R N A such as attachment of three prime c c a trinucleotide unit and modification of certain bases takes place will certain bacteria and almost all the eukaryotes.

### (Refer Slide Time: 40:10)



### (Refer Slide Time: 40:27)



All t R N A's have a common secondary structure represented by a clover leaf having four base prior distance the entire codon loop recognizes the corresponding m R N A codon, while the acceptor stem at the suitable amino acid to the growing polypeptide chain. The r R N A is central component of ribosome involved in protein synthesis in all the living cells. Prokaryotic seventy s ribosome is composed of 50 s and 30 s subunits, where s is a measure of the rate of sedimentation of respective components in a centrifuge, r R N A is that derived from longer precursors known as pre R N A. A single 30 s r R N A precursor is processed by several enzymes to give rise to 16 s, 23 s and 5 s, r R N A's in bacteria. In eukaryotes the eukaryotic 80 s ribosome is composed of 60 s and 40 s subunits.

(Refer Slide Time: 41:10)



(Refer Slide Time: 41:18)



### (Refer Slide Time: 41:55)

|            |            |                      |     | 75 ÷ |
|------------|------------|----------------------|-----|------|
| Eukaryotid | : ribosome |                      |     |      |
|            |            |                      |     |      |
|            |            |                      |     |      |
|            | 185        | 5.8S<br>Mature rRNAs | 285 |      |
| NPTEL      |            |                      |     |      |

## (Refer Slide Time: 42:21)



Where s is measure of rate of sedimentation of respective components in a centrifuge in eukaryotic vertebrates, a single 45 s r R N A precursor is processed by several enzymes to give rise to 18 s 5.8 s and 20 s r R N A's. The different classes of R N A basically, three classes messenger R N A, m R N A, which is the least abandoned around 5 percent of total R N A, and it provides a template for the protein synthesis or translation process. Transfer R N A or t R N A it carries amino acids in an activated form to the ribosome, which is in the intermediate abundance. And ribosomal R N A or r R N A, which is measure component of ribosome and provides catalytic and structural roles.

### (Refer Slide Time: 43:02)



So, what is function of different classes of R N A in protein synthesis or translation process. Messenger R N A, it is along sequence of nucleotides that serves as a template for protein synthesis. It is transcribed from a D N A template by R N A polymerase and gets translated into amino acid sequence of corresponding protein. The eukaryotic m R N A, requires extensive processing to form the matured m R N A, while prokaryotic m R N A does not require. The t R N A that is a relatively small R N A molecule involved in the protein synthesis process it binds to an amino acid at one end and base pairs with an m R N A codon at the other end. Therefore, it serves as an adapter molecule that translates an m R N A code into a sequence of amino acids.

#### (Refer Slide Time: 45:25)



r R N A it is the central component of ribosome's and it has both catalytic and structural role in protein synthesis. The ribosome's that houses this r R N A consist of large and small subunits, shall I discuss how protein synthesis occurs in following animation. Initiation of protein synthesis is carried out by binding of m R N A to a small ribosomal subunit such, that it is initiation codon most often a u g sequence it is aligned at the p side. The initiated t R N A that carried a modified methionine amino acid on it is acceptor stem, then binds to the ribosomal subunit by means, of codon entire codon interactions. The large subunit is then assembled on top of this to form the initiation complex.

#### (Refer Slide Time: 47:00)



The other initiation factors are also involved, which ensure correct positioning of all the components. The next incoming amino acid t R N A carrying the amino acid corresponding to the next codon occupies the a site. A peptide bond is then formed between the amino acid in a site and the p site with p site amino acid being transferred to the a site. The unbound t R N A, then leave the p site and it is moved to the exit or e site before being removed. When ten a tight bond has been formed the ribosome moves one codon towards the three prime end of the m R N A, such that the t R N A in the a site now occupies the p site and the a site is again free for the next incoming amino acid t R N A.

Multiple such rounds of elongation followed by translocation of t R N A's are carried out to form the growing polypeptide chain, then the ribosome encounters the termination sequence typically u a a, u a g, u g a. A release factor binds to the vacant a site and the polypeptide chain is hydrolised and released.

### (Refer Slide Time: 47:50)



### (Refer Slide Time: 48:03)



Other termination factors also aid this process once synthesis is complete the ribosomal subunits dissociate from each other. And all components are separated until commencement of next round of translation. So, in summary, today we talked about the importance of central dogma we discussed about the basics of D N A structure and function we then talked about basics of R N A structure and function. We discussed transcription and translation process briefly. We did not have enough time to go through the proteins, but in the next lecture, we will continue on amino acids and different levels of protein structure in more detail. Thank you.