Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Aritri Bir

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Week 01

Lecture 01: Foundation of Molecular Biology

Namaskar. So, I welcome you all to our new NPTEL online certification course. The course title as Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. This is a 12 weeks course as I have already told you and with me I have my co instructor Professor Arindam Ghosh and myself Aritri Bir. So, I am starting the very first class of this NPTEL lecture series that is the foundation of molecular biology. In this course we are going to discuss a lot about molecular diagnostics.

For today's section the concepts which we are going to discuss at the very basics of the molecular diagnostics or rather molecular biology where we are going to discuss the structure and characteristics of the two very important molecule that is DNA and RNA and of course, the central dogma associated with it. So, let us see what actually molecular diagnostics is. Molecular diagnostics is a branch of laboratory medicine which is constantly evolving and rapidly advancing in the field and it includes a lot of techniques which can analyze the biological markers at molecular level. What are those biological markers that is DNA, RNA, different proteins those are the molecular markers by which assessing them can help a lot in the diagnostics.

So, coming to the very early DNA discoveries, DNA is discovered in 1869 by Swiss researcher Friedrich Messner. Now, Dr. Messner was basically working on the composition of lymphoid cells where he by mistake has discovered a new molecule which is known as nuclein. Nuclein is actually the DNA associated with different proteins from the cell nucleus. After that there are a lot of studies which have been done and the pioneers are you all know James Watson and Francis Creek.

They have elucidated the double helical structure of DNA in 1953. Subsequently the work of Rosalind Franklin and Maurice Wilkins also contributed to the understanding of DNA structure. So, this pool of work has finally, given a clear concept of genetic information stored in DNA and their transmission pattern based on which we came to

know that this is the genetic basis of a multitude of diseases and development of techniques as required to detect and analyze these molecules DNA, RNA proteins and those techniques later became the key components in molecular diagnostics. Next is polymerase chain reaction is another milestone in the diagnosis in the discovery of molecular diagnostics that has been invented by Carey Muehlis in the year 1983. Now, PCR is basically the method where a minute amount of DNA is multiplied or amplified in an exponential pattern.

So, what we are getting millions of copies of a same specific DNA segment. Now, what is the utility of this in a very minute sample from a very small amount of sample where there is a minute amount of DNA that can be increased in amount amplified in amount which is the very basic steps in molecular diagnostics. In fact, PCR is considered as the first step of genetic analysis in majority of the techniques and this has been utilized profoundly in the field of molecular diagnostics. Next human genome project a very well discussed topic the project started in 1990 and is completed in 2003 and this is the work which has been done collaboratively by multiple international institutes they aimed to sequence or map all the entire genes of human genome. So, what HGP actually is providing it is providing the entire mapping of the genetic information or human genome present in our body based on that there are multiple helpful information we are getting like genes which are associated with different diseases.

The variations of genetic pattern which is actually associated with associated with mutation or various disease phenomena. Then also there are different targeted diagnostic tests and therapies that can be rather have already been designed based on the human genome project. Then CRISPR-Cas9 advisement this is the advancement this is the latest of 2020 because you know Emmanuelle Charpentier and Jennifer Doudna they got the they have been awarded Nobel prize in the field of chemistry for discovering this extremely sharp and helpful gene editing technique that is CRISPR-Cas9 clustered regularly interspaced short palindromic repeats as the CRISPR and associated with is the CRISPR associated protein 9 or Cas9. Now, this is definitely a genome editing technology which helps in precise modification of a very specific DNA sequence. Based on that based on this CRISPR-Cas9 multiple methods multiple diagnostic test can be developed it has a potential to correct different mutations and can design targeted therapies naming the gene therapy or the different novel treatment to treat different genetic disorder.

So, these are the different landmarks or milestones in the field of molecular diagnostics more we proceed in the coming 12 weeks more we will get to know about the different other techniques and associated discoveries along with it. Now coming to the very important keywords based on which our whole lecture series is actually depending upon these things. So, DNA or deoxyribonucleic acid this is the molecule which carries all the

genetic information of a being starting from prokaryotes to eukaryotes every in every organism this is the this is the portal or the carrier which carries the instruction for everything starting from the development functioning growth and reproduction of all the living organisms. Then gene what is gene? Gene is basically a part of the DNA or that part of the DNA which encodes the primary sequence of some final gene product a product which is getting synthesized from gene that can be the polypeptide or protein or can be the RNA with a structural catalytic function. So, these are the final targeted molecules which do the function in our body.

Genome is the complete set of the organisms gene on genetic material. Allele so, gene can have different versions or variations that can give rise to different traits. So, these are the alleles or the different variation of the gene and that can be dominant or recessive in terms of expression. Then chromosome so, chromosome is basically a long thread like structure found in nucleus that carries the genetic information and you all know there are 23 pairs of gene in human. So, these are the very common terms you need to know.

Then I am coming to central dogma. Now what is central dogma? Central dogma states that genetic information can flow unidirectionally from DNA through RNA to protein. Protein which is performing all the functions in our body. Now as per central dogma DNA gives rise to RNA which carries the same information as the DNA is carrying then that information is translated in protein. So, this is the central dogma.

Now there are variation that you will get to know. Then DNA or deoxyribonucleic acid by the name it is predominant that this is a polyribonucleic acid or poly deoxyribonucleic acid. So, this is made up of molecules which is known as nucleotides multiple nucleotides are there and each nucleotide contains a sugar, a phosphate and nitrogen bases. Now what are those nitrogen bases? There are two types one is purine another is pyrimidine. Now pyrimidine bases are cytosine thymine and uracil.

Remember uracil is present in RNA and it is counterpart which is present in DNA is the thymine. Then the purine bases adenine and guanine which are same in DNA as well as RNA. Now these nitrogen bases you can see these are the nitrogen bases which are attached to ribose or deoxyribose sugar ribose is present in RNA deoxyribose sugar instead of ribose is present in DNA. Now these sugar and fast sugar and nitrogen bases they form the nucleoside it is S ok nucleoside and this nucleoside when is attached to phosphate it forms the nucleotide fine. So, nucleotide contains all the three nitrogenous bases sugars and phosphate residues.

Now these nucleotides are attached to each other via different bonds in same strand as well as in the opposite strand. So, let us see what it is saying or the base pairing rule or binding. So, basically DNA is the double stranded helix there are two strands and they

are running anti parallel to each other. Anti parallel means here you can see one strand is running from 5 prime to 3 prime end whereas, other strand is just running to the opposite 5 prime to 3 prime end. So, these two strands are anti parallel to each other and they are spirally twisted upon each other.

So, what is created you can see there are a spiral twisting of these two bond which creates two types of groups one is major group another is minor group. So, this one is our major group and this one is our minor group. Similarly this one is our major group and this is the minor group. Now in a strand nucleotides are held together by phosphodiester bonds. So, these are our phosphodiester bonds.

So, here you can see these two nucleotides they are forming the phosphodiester bond and those phosphodiester bonds are formed between the 5 carbon of 1 nucleotide and the 3 carbon of the adjacent nucleotide. So, here you can see this 5 carbon of 1 nucleotide is forming the phosphodiester bond with this 3 prime carbon of the adjacent nucleostime nucleotide. So, this is happening in one strand whereas, the two strands are also held together by another type of bond that is weak hydrogen bond and that bond is formed between the nitrogenous bases fine. So, here you can see these are the bases here you are getting the sugar and the phosphate residue and the bases are here bases are taking part in forming hydrogen bonds. Now these hydrogen bonding form the base pairing follows the base pairing rule as stated by Sharghoffs.

Now Sharghoffs stated that is in fact, it found that the cytosine and guanine they are forming 3 hydrogen bond whereas, adenine and thymine they are forming 2 double bond. So, this is the base pairing rule which is followed through entire sequence of the nucleotide nitrogenous base pairing or in the DNA double helix where the cytosine is forming 3 hydrogen bond with the guanine whereas, adenine is forming 2 hydrogen bonds with thymine and that is our base pairing rule. Then in DNA most of the DNA double helices are twisted in a right handed pattern whereas, only one form that is Z DNA is a left handed DNA. Now as discussed in the previous slide the backbone of this double helical structure this is the backbone of this double helical structure it is formed by sugar and phosphate whereas, the bases are perpendicular to the helices you can see it is been if this is the strand the bases are forming like this. So, it is designed in a perpendicular or planar fashion.

Now sugars nearly at right angle and backbone has a negative charge because of the phosphate residues. Now the outer edges of this nitrogenous bases because they are projected outside these outer edges are exposed and available for potential hydrogen bonding also here different molecules can bind. So, base stacking hydrophobic interaction between adjacent base pairs they also contribute to the stability of the DNA along with the base pairing and also the phosphodiester bond. Now coming to the very

structure of the double stranded DNA. So, you all by this time you know purine pyramidine pairs with each other and this pairing contributes to the thickness of the DNA which is around 20 angstrom or 2 nanometer.

Now because these 2 strands are spirally twisted upon each other one turn of 360 degree of this spiral contains about 10 nucleotides on each strand. So, one spiral of DNA is so, here you can see this is one complete spiral of DNA which is around 34 angstrom or 3.4 nanometer. So, this is the pitch and by the rule of replication which we are going to discuss in next classes that the replication is a semi conservative replication here one strand acts as a template which gives rise to another daughter strand upon which a new strand is basically synthesized on replication. Then proteins also can interact with DNA at the exposed bases in different groups.

So, the major group which is 1.2 nanometer and the minor groups which is around 0.6 nanometer different bases can different proteins can bind and these are parallel to the phosphodiester back backbone and wind along the molecule. So, this is the very basic structure of DNA. Now along with the DNA there is a protein very important protein that is histone protein which actually helps the DNA to form a very compact structure the whole length of DNA is basically compacted over these histone proteins to get accumulated in the very small cell.

So, this histone protein has a very unique nature this is one very highly basic protein. Proteins are mostly of acidic nature, but this is one example of basic protein where huge amount of basic amino acids are present. There are 5 families of histone proteins H 1, H 2 A, H 2 B, H 3 and H 4. Now amongst them H 2 A, H 2 B, H 2 H 3 and H 4 they are around them basically these 4 proteins around them the DNA wraps and form a structure which is known as nucleosome. So, this is the structure of nucleosome here the core histones that is H 2 A, H 2 B which are lysine rich and H 3 and H 4 those are arginine rich lysine and arginine are the basic amino acids they form octamer.

So, each histones are present in 2 residues. So, that is one octameric form and it is arranged in handshake motifs arrangement. Now around this octamer 1, 2, 3, 4, 5, 6, 7, 8 around this octamer DNA of around 146 base pair is wrapped for 1.6 times 1.65 times in a left handled super helical term and that is forming the nucleosome.

So, this structure is our nucleosome and it gives rise to a bit on string formation. I am showing the next in the Bison strings of formation in the next slide. And then there is H 1 histone protein it is loosely associated with the DNA. Now let us see how these structures are formed. So, this is the higher organization of DNA where you can see this is the double helical structure of DNA which is wrapping around the nucleosome or the octameric histone proteins and giving rise to this beads on string appearance.

So, it is just like this. So, here are the octameric protein. So, this is how the Bison string appearance is formed. So, the octameric nucleosome now has one H 1 histone as well which is loosely attached. So, you can see between 2 nucleosome there is a string or a strand of DNA.

Now this DNA is known as linker DNA. So, once again so, this is our linker DNA between 2 nucleosome and H 1 histone is basically the gateway of this nucleosome. So, basically it is locking the structure of nucleosome or the entry. So, if you consider this is the histone protein the entry of DNA here is the histone H 1 histone and after wrapping when the DNA is going out again it is passing through this gate. So, H 1 histone is basically locking the DNA over the nucleosome. Then this nucleosome fold up to form a 30 nanometer fiber here you can see this is the 30 nanometer fiber which has been which is the folded pattern of nucleosome forming the chromatin fiber.

This chromatin fiber once again is forming loops for 300 nanometer in length this is the loops which again is now compressed and folded to produce a 200 nanometer white fiber. And this white fiber finally, is coiled up tightly to form the chromosome. So, this is the higher organization of DNA. Next a very important characteristics of DNA that is denaturation the 2 strands of DNA can be separated or melted by heating. Now, T m is the melting temperature where half of the strands of DNA is basically denatured or the DNA is opened the 2 strands of DNA are opened up.

Now, because there are 3 hydrogen bond between G and C and 2 hydrogen bond between A and T it needs more energy to denature 3 hydrogen bonds. So, the regions which are rich in GC they need more temperature. So, melting occurs at higher temperature. Now, melted strands can once again be re associated if the temperature lows down this is known as annealing of the strands. Formamide is one such compound which lowers this temperature by disrupting the hydrogen bond.

So, any chemicals which opens up the internal hydrogen bonds lowers the melting temperature the required energy is less because there is less hydrogen bonds to dissolve. Phosphodiester bond in this denaturation is basically not broken. And one very important characteristic that is hyperchromicity of denaturation where optical absorbance on denaturation of DNA is basically increased that is known as hyperchromicity of the DNA. Coming to supercoiling now linear DNA is twisted around its own axis and forms the supercoils. And this supercoils gives strain over the DNA strands.

Now, DNA strands can exhibit positive supercoil or can exhibit negative supercoil based on more or less tightly on DNA strands. Now, this degree of supercoiling strand has impact over the functioning or biological process because more compacted DNA is

less available to bind with different proteins or to perform different function. Topoisomerases is one such enzyme one very important enzyme that can introduce changes in this supercoil. They can introduce negative supercoil they can introduce positive supercoil as well. There are two types of topoisomerases and gyrase or DNA gyrase is type 2 topoisomerases.

Now, why I am saying about this discussing this enzymes this will be required in your next few classes where we are going to discuss replication. Then introns and exons now only 10 percent of the human DNA basically contains gene the rest are actually silent areas and in that about 90 percent of the DNA are permanently inactive. Now, a segment of the gene which codes for protein is known as exons or expressed region whereas, the intervening region between this expressed regions or exons are known as introns which do not give rise to which are not coded takes part in coding the proteins. Now, when from DNA through mRNA to protein is formed remember this mRNA contains this primary transcripts of mRNA contains exons as well as introns and these introns are cut off by a process known as splicing.

So, if there is exon and intron as well. So, what happens this introns in cut off and these two exons are joined and that is our splicing. This process we are going to discuss in next classes. So, introns are basically not translated to give rise to protein. Now, coming to RNA so, RNA the structures are very similar to DNA, but there are few differences like RNA is the single stranded one except there are few hairpin structures are there which are double stranded instead of deoxyribose sugar there is ribose present thymine which is present in DNA instead of this thymine there is uracil except in t RNA where thymine can be seen and also because there are not there are only one strand Sharghov's rules are not required. So, there are different types of RNA messenger RNA transfer RNA and ribosomal RNA.

So, mRNA t RNA and r RNA apart from that there are other types of RNA as well like HN RNA heterogeneous nuclear RNA SN RNA small nuclear RNA MIRNA that is micro RNA and SIRNA small interfering RNAs. So, these RNAs we are going to discuss in those segments where we are basically discussing their functions as well as their molecular role in molecular diagnostics. Currently in this class I am going to discuss these three. So, structure of mRNA now mRNA is the information carrying molecule and it consists only 5 percent of the RNA in cell is most heterogeneous in size as well as in base sequences by name it is evident that it carries the messages or the information from DNA to protein from gene to protein. Now the sequences of nucleotide in mRNA is basically complementary to the sequence of the nucleotides in templates.

So, if the DNA strand contains like these the RNA strand will be complementary only instead of t their will be u uracil. So, this is the base complement base pairing

complementarity between all the nucleotide strands let it be DNA or DNA RNA like that. The sequences which are carried on mRNA is read in the form of codon that is three consecutive nucleotides they form the codon along with that mRNA contains non translated region or non coding sequences and also coding region which codes for the protein. So, here is the coding sequence in the mRNA molecule. Now apart from that mRNA also contains a cap at the 5 prime end and a tail at the 3 prime end the cap is composed of 7 methyl guanosine triphosphate cap the functions are basically giving protection from the 5 prime exonucleases.

In case of this 3 prime tail it also confers protection for the 3 prime exonucleases. This 3 prime tail is basically composed of multiple adenine residues. So, this is a poly a tail for around 20 to 250 adenine residues. So, this is the structure of mRNA. Now coming to the tRNA or transfer RNA which mostly composed of 74 to 95 nucleotide residues and tRNA basically comes as a precursor molecule remember mRNA also comes as a precursor molecule and there is a nuclear processing of this precursor molecule to form the tRNA.

Now function is in the name transfer of the amino acid from the cytoplasm to the protein synthesizing machinery that is sRNA or soluble RNA. So, it is basically acting as a adapter between the amino acids and the mRNA code. So, there are at least 20 species of tRNA which are carrying these amino acids from the cytoplasmic pool to mRNA during protein synthesis. Now tRNA has primary structure as well as the secondary structure. Primary structure is nothing, but the nucleotide linear nucleotide sequences which allows extensive interest and complementarity whereas, secondary structure is basically this.

So, each single tRNA shows extensive internal base bearing and finally, it is forming such bonds upon its own strand that gives rise to a clover leaf structure and this structure is basically stabilized by the hydrogen bonding between different bases. Now if we describe this secondary structure there are different arms or loops in this secondary structure starting from acceptor arm, anticodon arm, d arm, t-psi-c arm and extra arm. So, this is the amino acid arm or acceptor arm which accepts amino acid from the cytoplasm. It is it has 7 base pairs and the end sequence is cytosine, cytosine and adenine this CCA.

Now the CCA attachment is done post transcription. Initially when the tRNA is formed from DNA there is no CCA, but after transcription this CCA is attached and in this CCA basically for this adenine the 3 prime hydroxyl of this adenine form the bond with the carboxyl group of amino acid with the enzyme aminoacyl tRNA synthetics. Coming to the anticodon arm which is located just opposite to the amino acid arm that is the anticodon arm. Now anticodon arm contains 5 base pairs and it is recognizing the triplet codon present in mRNA. So, in mRNA if it is like this the code of tRNA is like this.

This is basically you this. So, this is how tRNA binds with its anticodon arm to the codon present over mRNA. This is the mRNA that is why it is known as anticodon arm because it has the base complementarity with the mRNA codons. Next coming to d arm contains 3 to 4 base pairs and recognizes a site for the enzyme aminoacyl tRNA synthetics. Next is t-cycium which is just opposite to the d arm this is our t-cycium where pseudo uridine is present and it is helping the tRNA to bind with the ribosomes the protein synthesizing machinery. Then extra arm or variable arm variable arm because it is not always present in all types of tRNA.

When present it is about 3 to 5 base pairs. Now tRNA has different classes class 1 contains this variable arm. Now this is the long extra arm which is 13 to 21 base pair in length. Now these variable arm is basically variably present on in different tRNA, but the other 4 arms are persistent or common for all the arms this is the this is the variable arm. So, you can assume how mRNA over which tRNA is bound by this anticodon arm and in this acceptor region or acceptor arm there is attached amino acid. This is how it acts as the adapter between the amino acid and the mRNA and that is how central dogma or the genetic information is flowed from DNA via transcription to form the mRNA.

Also in transcription process rRNA and tRNA synthesized then from mRNA protein is synthesized via translation. So, this is how the central dogma maintains the informational integrity in all the living organisms. So, these are my references. Thank you and see you in the next upcoming class.