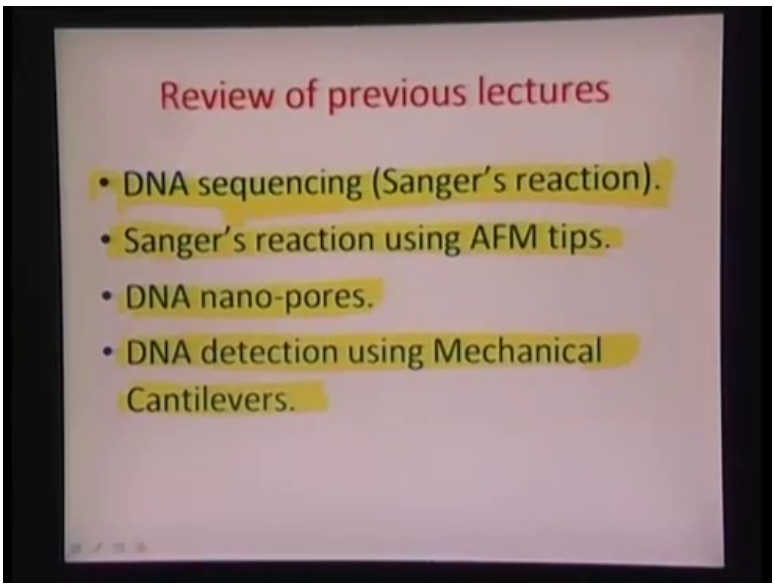
BioMEMS and Microfluidics
Prof. Dr. Shantanu Bhattacharya
Department of Mechanical Engineering
Indian Institute of Technology, Kanpur

Lecture - 22

Hello and welcome back to this lecture 22 of Bio Micro Electro Mechanical Systems.

(Refer Slide Time: 00:14)



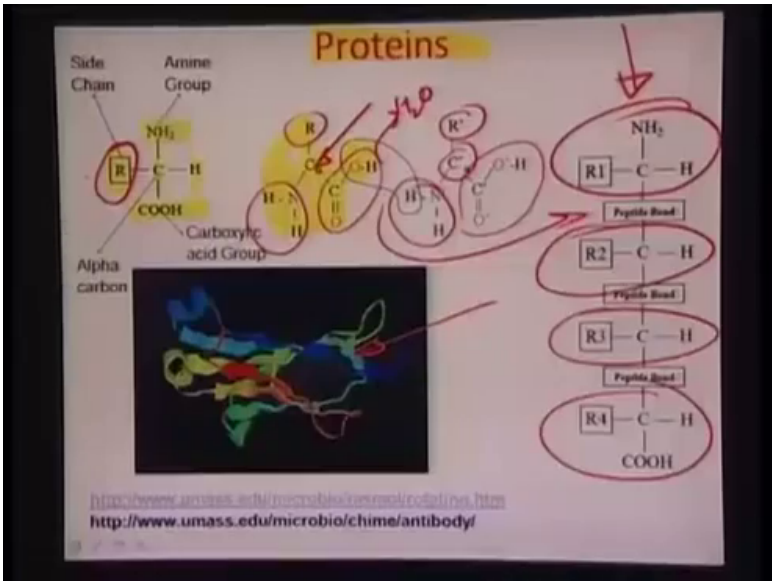
Let us quickly review what we did last we talked about DNA sequencing using William Sanger's process, where in we actually have a chain terminated ddNTP set of groups and four different reactions as you know the dd or the dideoxyNTP is essential for terminating the chain of the process of replication. So, you can have various lengths terminated long different specific groups. So, when you run all these products of all the four reactions along a gel parallel to each other in four parallel tracks, you get different lengths which identifies what really is the particular group at that instance or at that length from the overall read out.

So, essentially you can actually sequence of DNA by looking at, where along it is chain the ddNTP whether it is an ATP or CTP or GTP or thymine triphosphate it is able to cleave and terminate the chain replication reaction. We also solid the glimpse of the technologies, the most modern evolving technologies where this Sanger's reaction can be used and the transport of DNA could be done on a molecular level using STEM pen or AFM tip. we further try to explore DNA Nano pores, where in you basically see how in nature, you know self-assemble proteins over lipid bylayers can be used as DNA channels, DNA Nano channels.

Similarly, you could also use silicon on insulator platform to create a thin Nano pore over this platform and then use it successfully to translate DNA from rich to a lean side using and electrophoresis potential. So, this can also be an identification protocol for sequencing DNA as we talked about in details. We discussed some of the other alternate strategies of detection, especially mechanical detection whenever you talk about cantilevers, you know mechanical cantilever is essential we have diving board like a structures.

So, you basically have a change in surface energy causing a deflection delta z which is given by Stoney's equation and then from that you can actually find out whether what is the change in surface stress causing a certain delta z value which you can detect; otherwise, using optical or other means. So, therefore, this kind of summarize what we did last time I would like to now look at proteins, so we will do some primary introduction of proteins today and then we also like to correlate how this fantastic Nano machinery of a living cell kind of translates the information, which is available as a compressed DNA or RNA into an amino acids sequences or proteins. So, let us look at what protein really is.

(Refer Slide Time: 03:21)



So, if you look at this figure right here, a protein really is a bunch of different amino acids. So, what is the amino acid, you have alpha carbon, you have a carboxylic group here, you have a side chain which can be a polar or non-polar moiety and then you have a amine group at one end and then there is a hydrogen on this alpha carbon. So, this essentially is what an amino acid would look like, it is a carboxylic group on one side and NH2 group another side of an alpha carbon that is what the basic idea of an amino acids.

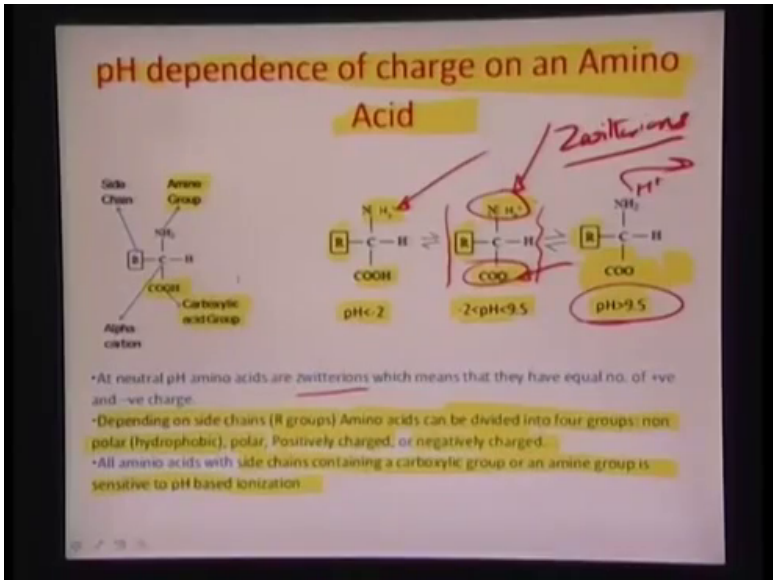
Now, the way that this amino acids kind of bond together is true this amide bonds. So, you have this let say is one of the amino acids with the C alpha, the alpha carbon right here and then this is the amine group and this is the carboxylic group essentially the CO, H is a carboxylic group on the acid and the other amide here and the amino acid here also has another group let say R dash on the side chain, here it is R and this alpha carbon is connected to and NH 2 and also COH. So, there is a bondage between the OH on the COH and NH 2 here thus producing a water molecule and you have what you called a peptide bond, this is essentially is something or peptide bond C single bond n and this double bond on one side with though on the N side you have a single bond H, so that is what essentially a peptide bonds.

So, several such amino acids bond to each other with different 10 groups may be with this peptide bonds to forms as says arrays of a change of molecules call proteins. So, protein is basically defined as a chain of amino acid inter connected using this peptide linkage. So, this can be essentially a model of how protein would typically look like when in solution of a certain kind.

So, one important point about protein is that it undergoes rapid conformational change, because of this R group. The R group as you see here can be polar or non-polar depending on the type of you know applications that are if depending on the type of various amino acid groups that are present in this world. So, there are above 20 different such amino acid groups, which are categorized into polar, non-polar, ionic, non-ionic, etcetera.

(Refer Slide Time: 06:01)

So, essentially let us look at some of the interesting things and the pH dependents of such an amino acid. So, if you really look at this carboxylic moiety here, it is a COOH, so an acid carboxylic acid group and you have an amine group. So, a protein when put in a certain pH can absorb and excess hydrogen ion onto it is NH forming NH 2 forming NH 3 plus or it can depending on if the pH's like basic generate hydrogen ion thus creating COO minus. What is interesting here is really that it is a function of pH. So, suppose the pH is very, very acidic it is less than let say two this hydrogen ion concentration.



So, in this instance as you see here, because of the excess available hydrogen ion in this solution, there is a tendency of this amine group NH 2 to protonate to get into NH 3 plus that is one aspect of how a positive charge can be produced essentially on the amino acid. Now, there is also a range of pH let say in this particular case here it is between minus 2 and 9.5. So, this form slightly acidic to kind of the reasonably high basicity in this particular pH range, there is a tendency of the carboxylic group, the carboxylic group to protonate. So, because there is an excess of OH minus concentration. So, therefore, there is a tendency of these COOH here to protonate and get converted into COO minus as you can see here in this case.

Now, in this kind of a condition also the total amount of positive and negative charges on this particular moiety are equal and opposite of each other. So, they are essentially known as zwitterions, there is a term to denote these kind of moiety is where there are ion centers within the same molecule, but then the amount of positive and negative charges are equal and opposite and so they are called zwitterions. So, they have equal number of positive and negative charges.

But, if you change the pH into more basic, let say the pH goes above 9.5 in that case there is a tendency of this the excess hydrogen ion which was already formulated here in zwitterions of the state to kind of protonate into the solution. So, the H plus goes out into this solution, because the basicity simply is high enough for any positive ion or in positive charge, hydrogen moiety on the amine group to be present, so it kind of protonates.

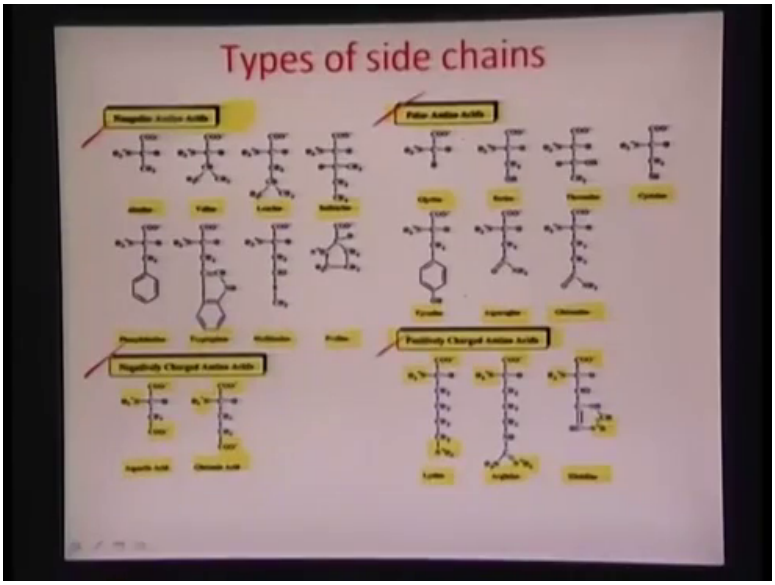
So, here again as you see the overall charge of the amino acid reduces to negative. So, the charges state of an amino acid really depends on the pH. So, if it is very highly acidic you

have a positive charge on the acid, if it is in a certain range were the equal amount of positive and negative charges, the amino acid essentially behaves this zwitterions and then if the pH is carried on from their two more basic that is the tendency of the protonation to happen again and then you have essentially a negative charge on the particular amino acid.

So, depending on the sides change of this R group here, this is very, very critical the R group maybe polar, maybe non polar depending on whatever it is or whatever it is nature is there is a tendency of the proteins to change their shapes and sizes by folding. And essentially this is also known as a confirmation change, it is technically known by that name, but the way that proteins formulates into a closed loop or a structure is really a function of what are they are on the R groups on the protein or these site changes of proteins are very, very critical to determining the interaction of the protein in a aqueous or a non-aqueous environment rounded.

So, amino acids can you know depending on the side change between divided into four groups non polar or hydrophobic polar or with the same amount of you know this kind of separated charges positively charged where there is an additional H plus ion on the particular amino acid or negatively charged where there is an additional minus charge on the carboxylic group by sending out H plus into the solution. So, there are essentially based on the R groups you can divide all the amino acids into four different groups. So, as we already learned that this side change containing carboxylic groups and amine as groups is really, really sensitive to pH based ionization that is one aspect.

(Refer Slide Time: 11:13)



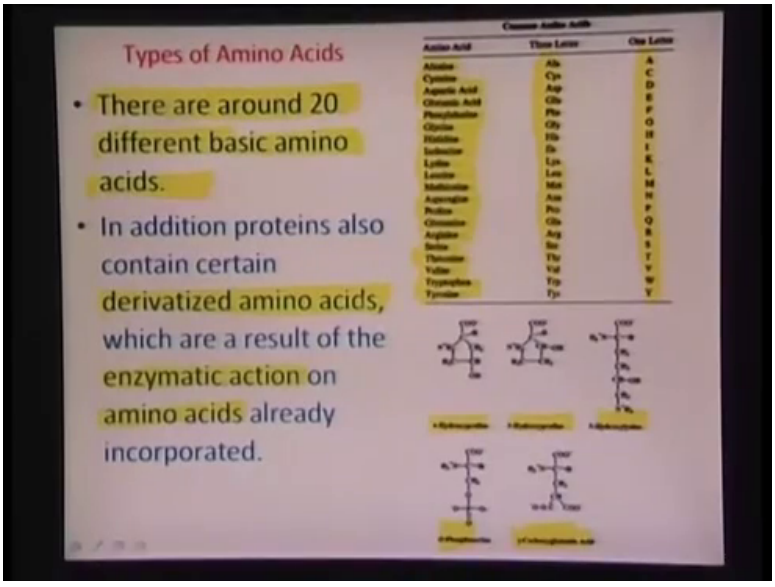
So, let us look at what are the R groups really and if you can categorized all the different amino acids they can essentially be drawn out as about 20 different amino acids. So, all the proteins that you see around the 1000 of chain or 1000 of types of chains are formulated from this 20 basic amino acid groups that are present.

So, therefore, you can say there you can categorized the side chain of such an amino acid into an non polar moiety like, alanine, valine, leucine, isoleucine, phenylalanine, tryptophan, methionine proline. So, these are essentially the non-polar from the amino acid, the R groups on the amino acids, then you can also categorize the R group into polar amino acids, where this R has a kind of slight shifting of the positive and negative charge centers although overall molecule is neutral, but just as it happens in water that there is a slightly more electro negativity in one center of the molecule and depending on which there is a charge separation, here also there is a kind of charge separation within the same molecule. So, that is why polar.

So, the R groups can be having a you know different names like glycine, you have serine, you have threonine, and cysteine, and tyrosine, asparagine and glutamine, and these are some of the polar amino acid groups. You could also divide the R groups into negatively charged amino acids and they can be like aspartic acid or glutamic acid like you can see these two cases and then you have also the positive. So, here in the negative charged amino acid is you see there is an excess amount of negative charge centers over really the positive charge center, here also it has excess access amount of negative charges in addition to the only positive charge that you have here. So, these are essentially or function of pH states, various pH states, but there is of course, an additional negative charge here.

And then you have a categorize of positive fully charged amino acids like lysine, arginine and histidine, where you can see there is a negative charge center, but then essentially there are two amines which are a certain pH would be able to even in it is form in the true same pH range has two charge centers for positive charges and one really for the negative charge. So, these are categorize into positively charge amino acid. So, in a nutshell depending on this type of side chain is 20 different amino acids can be categorize into various kinds like non polar, polar negatively charged positively charged.

(Refer Slide Time: 14:05)

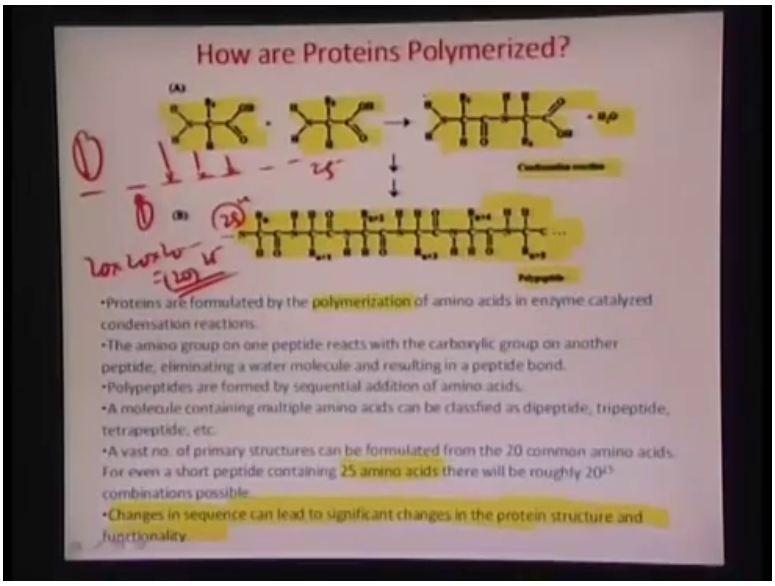


What is also important here is that let us just do some combinatorial to find out the kind of sequences of the total combination of sequences that are available to us which can define the whole protein world essentially. So, there are around 20 different amino acids. So, this is kind of a list of all these 20 different amino acids, there are polar, non-polar, negative or positive charges, you just saw in the last slide about these are some of the synonyms, which are used

either in three letters or one letter as you can see here.

And essentially the proteins also contains certain derivatized amino acids, which are result of enzymatic actions on amino acids already incorporate. So, you have because of some enzymatic actions on these principle amino acids there are certain derivatized forms, you know in addition to the 20 basic amino acids these all kind of represented here, but I am not going to go into this it is beyond this scope of course, to go into this enzyme derivatized amino acids essentially, but there are lot of these enzyme derivatized in addition to the 20 basic amino acids that we have.

(Refer Slide Time: 15:38)



So, if you want to do a combinatorial really let say we are talking about at 25 amino acids sequence and each of these 25 places can have about 20 different amino acids. So, if we can repeat a sequence then the total number of permutations that we will be having is about 20 to the power 25 moieties, you know such a vast amount of arrangements available only for a 25 amino acid sequence. So, you have 25 group like this and you can have 20 amino acids in each of these area. So, essentially 20 into 20 into 20 so on up to the 25'th terms. So, 20 to the power of 25 combinations, so this is the vast amount.

So, if you want to predict what is the one sequence out of these 20 to the power 25 sequence of amino acids is really a huge number to deal with. So, these are essentially the primary structures of the proteins that I am just going to come to that in a little bit what I would like to show here is that the way that proteins would actually be formulated or in a sense polymerization process, where in two amino acid groups are condensed together in enzyme

catalyzed condensation reaction, why we call it condensation is that where there is a displacement of a H2O moiety or you know a water molecule we call that reaction a condensation reaction.

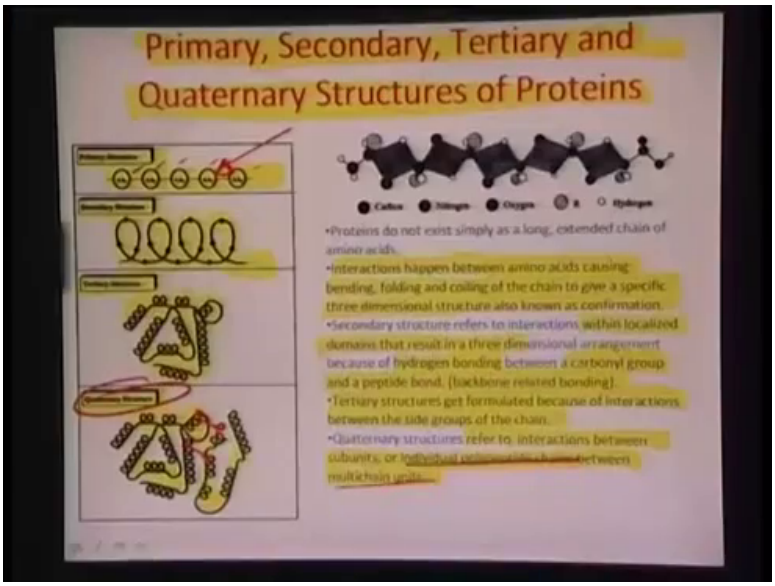
So, here this processes is enzyme catalyzed, so there is enzyme which would be able to formulate the peptide bonds between the two amino acids. Now, this is a very important concept which is very well applied to even nature, where Mother Nature tries to replicate the amino acid from MRN sequence in the translation process with they also use enzymes to cleave bounds on to make the peptide bonds. So, it is enzyme catalyzed condensation, we already know that H2O moiety goes out one on taking the OH from the carboxylic on one and H from the NH 2 an other, but this process happens on the presence often enzyme, so peptide bond is formulated.

So, here in this illustration it pretty much shows how this system. So, you have two amino acids sequences here and then there is an enzyme which would convert this and this condensation reaction. So, there is a H2O molecule, which is eliminated and you can have polypeptides change like these where in you have several of these peptide linkages and more than one amino acids, so proteins can be this large.

So, molecule therefore, depending on how many peptides molecule would have is classified either in to di peptide or try peptide, tetra peptide so on so forth and that also determines how many amino acids you have. So, one protein or one peptide linkages associated with two amino acids. So, you could find out the total length of the protein chain based on and categorizing in to a di or try or tetra peptide so on so forth.

So, we already know that how many combinatorial can be processed in possible only in a 25 amino acid sequence of protein this small sequence of protein can have about 20 to the power 25 combinations of if you can club all these possibilities of you know each place being felt by 20 different amino acids. So, changes in sequence of a protein can lead significant changes in the protein structure and the functionality. So, if the sequence is altered slightly it is where it is slightly you can have all together difference structure of the protein of interest.

(Refer Slide Time: 19:51)



What is also important here is that if we really categorize proteins into different kind of structures they can be classified principally into the primary, secondary, tertiary and quaternary structures and let us just investigate little bit about why they are called so or what are the difference? So, this cartoon here illustrates very nicely these four different structures. So, proteins do not really exist simply as a long extended chain of a amino acids it is hardly possible to exist in a natural state and a long chain.

One of the reasons why that is so is that because there are lot of OH groups and also there are lot of side chain interactions and there is also tendency of hydrogen on the NH 2 or COH the OH on the COH to interact with the solvent, where it is n and so there are additional hydrogen bonds which are created in between the molecule and between the molecule in the solvent in question. So, confirmation really is a very dynamic process the overall structures that happens is a very complicated structure just because of the several hydrogen based interactions between the protein moieties and it is surroundings or even molecules within the same protein moiety.

Now, primary structure those very important in an understanding what is the sequence of the amino acids on the proteins. So, essentially primary structure is the basic foundational structures it gives a kind of you know relationship between what follows what in terms of amino acids sequences, although it does not exist in nature, but it is always wonderful to represent protein with this primary structure to find out the amino acids sequences that it has you know.

So, interactions between of course, the amino acid change would cause the proteins to bend and fold. So, interactions happening between the amino acids cause bending, folding, coiling of the chain and eventually what emerges is a specific three dimensional structure this known as confirmation as have been discussing before. So, secondary structure as you can see here in this illustration is really formulated within localized domains. So, because of course, because of presence of hydrogen bonding and particularly because there is a carbonyl group and a peptide bond. So, in between the protein back bone itself there is a lot of bonding because of these two moieties. So, peptide bond on one side and then may be the carbonyl group on the other side.

So, secondary structure is like kind of coiling like this. So, you can say that if you leave this is a very fine analogy that I keep giving that you have a straight, you know let say hose pipe and this was otherwise coiled this hose pipe was otherwise in a coiled manner and coiling is by means of either stressing in a heated environment or something. Now, whenever you stretched totally you know intention and put it intention and stretched this is the primary structure that it could have, but the moment you release the two ends it is going to recoil back and because the stresses which was there would try to align it back in to the equilibrium and that is the secondary structure and when you drop this let say this is the telephone cord we are talking about.

So, now, you have a case where this you have just about released enough. So, that this structure gets coiled and then you drop the whole structure. So, what would happen is that this coiling would start to bend among itself and there would be an interaction between the coiling itself. So, that is what the tertiary structure of a protein would be looking like. So, there are interactions which are beyond the local domain, but there essentially confine to the same molecule, this is on the one single molecule that interactions between different domains of the molecule rather than just adjacent interactions as work happening in the secondary structure here. So, this is what the tertiary structure of a proteins.

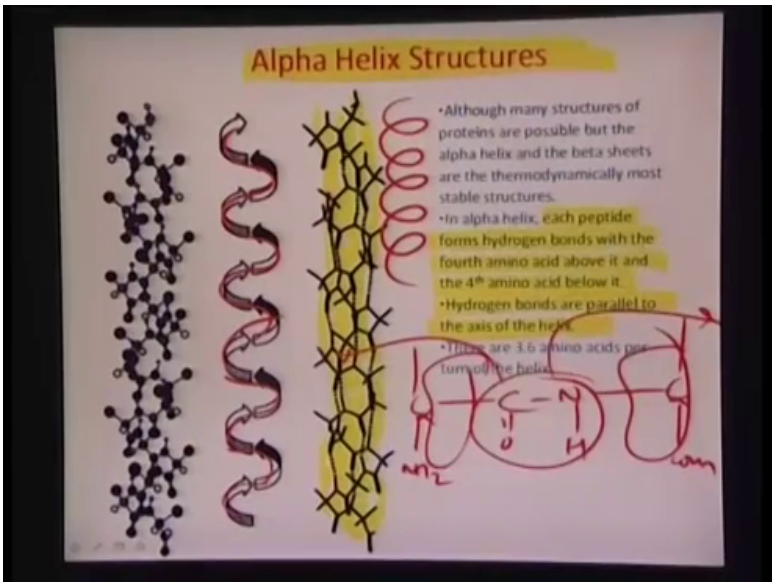
And then of course,... So, formulated because of interactions between side groups of the chain in different regions of the same molecule and then there is this quaternary structure, which is essentially referring to introductions between sub units or individual polypeptides chain between multi chains units. So, here for instance there is one protein which is present here in this area and another protein may be which is present independently in this area, so there is always going to be... So, there are two tertiary structures present close by. So, there are going to be interactions between the two structures.

So, therefore, you know between sub units probably of the same protein or individual polypeptide chains all late side by side in a solution those kind of multi chain interaction develop what you call the quaternary structure. So, therefore, starting from a simple to a very complex proteins have increased degrees of complexities in their confirmations or in their shapes and sizes.

So, if you look at the proteins in general all though there are many structures of these proteins which are possible we can principally classify the base units of these structures just as we have let say different you know units cells in a certain lattices we can also classify the proteins in terms of some base units repetition of which would essentially build the molecule you know as could in the crystal structure as you know the repetition of the units cell in the three dimensions makes the whole molecule.

So, there are two principle structures which are foundational for formulating or demonstrating the overall complexity in the protein structure and these so called an analogs of unit cells or known as the alpha helix and the beta pleats of proteins. So, you can either classifying to an alpha helix structure and I am just going to come in little bit into what alpha helix really means or there is a beta pleat where you can say that the protein is kind of rotating and the molecule is like big multiple access together compressed. So, there are parallel sheets of amino acids change which are running at different levels.

(Refer Slide Time: 26:53)



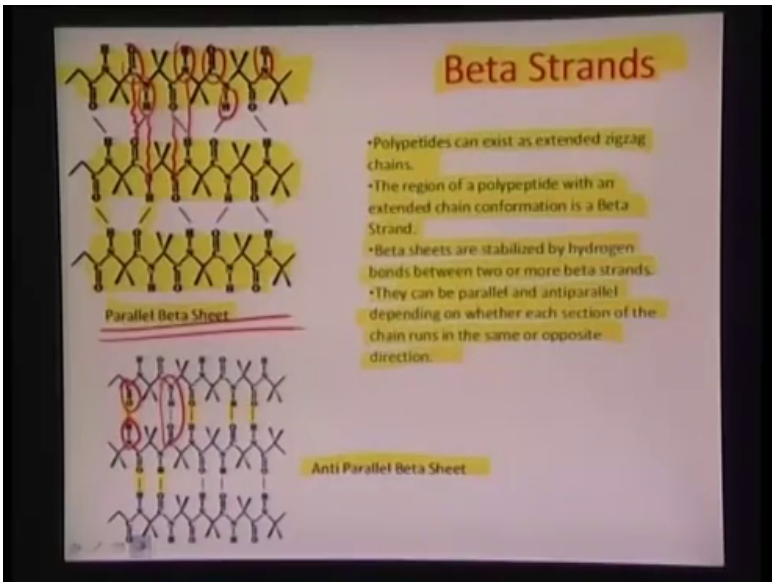
So, in alpha helix as you see right here, you can see an alpha helix here in this region, this is an alpha helix really. So, each peptide here from hydrogen bonds with the fourth amino acid above it and the fourth amino acid below it. So, the way it goes us that this peptide linkage which is essentially this C N double bond O and H here and then you have the other portion of the protein in both sides here and alpha carbon and you know things attached to it. So, here you have NH 2 attached, here you have a COH attached something like that.

So, this peptide bond with the fourth amino acid in the sequence ahead and the fourth amino acid in the sequence back their formulates hydrogen bonds, because of which there is a tendency of this molecule to kind of form alpha sort of structures and these is on a three dimensional basis that it is formulating the alpha. So, it is more like this that the alpha structure that we are talking about. So, essentially... So, it kind of coils just as any telephone cord would do and there are bondages form between the peptide and every fourth amino acid on either end. So, fourth amino acid in the bottom, fourth amino acid in the top they would have hydrogen bonds.

Essentially in one helical unit or in one unit of the helix that is formulated there about closed 3.6 amino acids in this kind of an orientation, hydrogen bonds would define alpha helix element, where in one helix you have only about containment of 3.6 amino acid. So, I mean there is no physical interpretation of the 0.6 only that 60 percent of one full amino acid is within the chain in additional to the three units that is how you define the helix length of this particular structure.

So, also one interesting factor here is that the hydrogen bonds are kind of parallel to the axis of the helix. So, that is the reason why this alpha kind of structure is formulated, this elliptical are this out say this helical kind of structure is formulated, because all the bonds is essentially in all the different areas of the proteins are parallel to this central access of the helix which confines the molecule with in this helical domain. So, that is what alpha structure is.

(Refer Slide Time: 29:50)



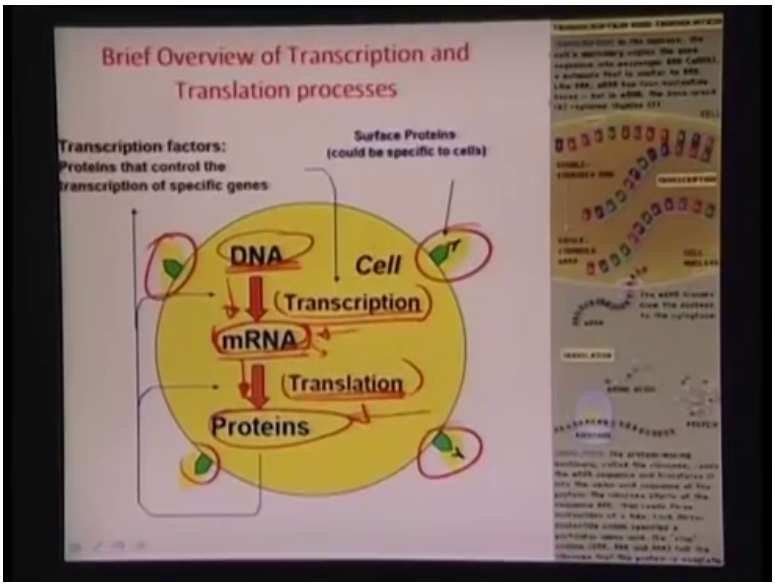
So, the second kind of like analog for unit cell you may call of the second kind of the structure that is most commonly use this units structure in protein of the beta stands, the way these beta stand structures can goes that they are like parallel sheets or they are like anti parallel sheets, there are two different kind of beta structure, where this proteins would be laid on the top of another like sheets of different amino acids sequence as you can see here, these are parallel because as you see you know in both the molecules here the groups are kind of to gather, this is the C double bond O, this is the NH, this again is C double bond O, this is again the NH. So, you have parallelity in the way this groups are placed on a single sheet of these beta sheeted structure. So, that is why parallel beta sheets, you may have an anti-parallel configuration, where the C double bond and one would comprise are would phase the NH one other another are vice versa.

So, essentially this is an anti-parallel kind of arrangement. So, therefore, polypeptides can extend as zigzag chains as you are seeing here call beta strands and the region of the polypeptide with extended chain the conformation is known is a beta strand, beta sheets is stabilize by hydrogen bond. So, as you are seen here there is a hydrogen bond here between the C double bond O on the carboxyl in the NH on the amine group, similarly there is another

one here between the C double bond O on the lower sheet on the NH on the upper sheet so on so forth. So, there are lot of bonds all though whether in parallel or in anti parallel of course, the anti-parallel has a lot more stability in the parallel has.

And so therefore, they can be parallel or anti parallel depending on whether is section of the chain once in the same or the opposite direction and the beta sheets as stabilized by hydrogen bonds between two are more beta strand, this is the anti-parallel beta sheet, this is the parallel beta sheet. So, in a nutshell these are the two structures alpha helix and the beta sheets.

(Refer Slide Time: 32:00)



Now, after going to this little bit understanding the mechanical aspects related to the molecule structure I would like to go into a very interesting area which is also known as DNA transcription and translation processes. Why it is needed is probably one of the most fascinating ball games today of chemist as well as molecular biologist, because this in a sense gives the cell it stability it is physiological stability. So, there is really it is like a genetic code it is like a small computer the cell which can convert a compressed code of this ATCG sequences into sequence amino acids or proteins and the proteins as we know are more or less present, you know the cell membrane and which is also responsible on the state of health the physiological state of the cell the way it goes replicates everything is related to the surface proteins that come on the surface of the cell.

And the process which allows these proteins to be generated is actuated deep within nucleus of the cell, where the messenger RNA is essentially formulated from section on the DNA, section on that small region in the chromatin in nucleus and there is a compressed sequence

of helical coils of DNA which we also known as chromosomes. So, that is a process with in the cell, where the DNA molecule gets converted into a messenger RNA molecule and this is the compression algorithm.

So, the amount of data you have on the DNA certainly much, much bigger and comparison to the amount of data you have on the m-RNA, it is a much shorter information sequence there on the DNA. So, you have a compression algorithm here. So, you have a messenger RNA and this process is known as transcription. So, you are making a transcript and you are not changing the language really, the language is same as ATCG only in this case T the thymine is replace by an uro cell as is normally the case with the all RNA moieties ribonucleotide moieties. And of course, this is a single stranded structure which is formulated from the DNA which is the double stranded structure.

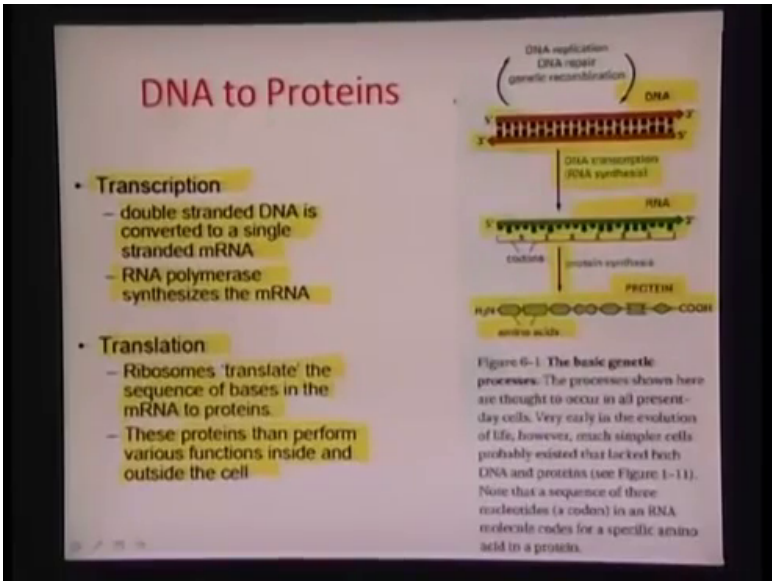
Now, the m-RNA gets generated within nucleus then it goes all the way up to outside the nucleus into applies called the endoplasmic reticulum it is an organelle which is their inside the cell and essentially called the protein where how. So, there are this intelligent Nano machines called ribosomes and the surface of this into plain this organ of the cell which essentially you take up a m-RNA sequence and then convert that into a protein sequence and the way it does is very fascinating there are some amino site transfer RNA's which are moving around in the solution and that is a group to group matching or a mismatching which Leads to the amino acids sequence as the m-RNA or the messenger RNA sequence is red at the nucleus level.

And thus amino acids sequence of the protein is formulated by this process on translation, why it is called translation is because the language of the m-RNA which was more in terms of ACG and U is now converting into altogether different language. So, you are translating the information an m-RNA into sequence of amino acids. So, it is a language conversion that is happening. So, that would result in you know development of the protein surface here and this is a function of how the cell would strain a certain environment.

So, the way that a cell reacts by having a signal from it is an environment would be in term expressed by or would be term kind of try to we back into... So, the cell will be try to be the equilibrium based on whatever response of stimulate is getting from the environment and the way that the equilibrium happens is through producing or protection of a set of proteins which would exactly cause the deviation from the equilibrium to be balance to back. So, that and the cell come is back to the equilibrium you know it is depends on what is the existing

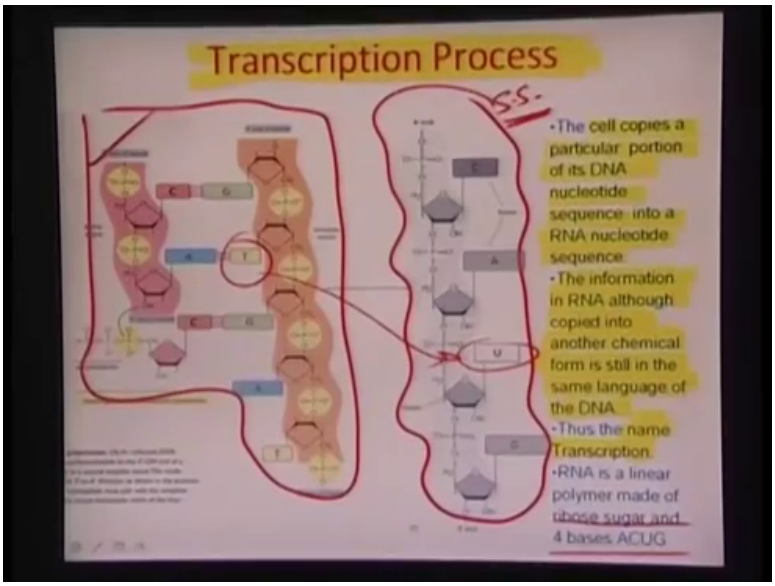
state physiological state and what is the deviation from that state and what would be needed to control that deviation from that state, these are the three steps which defines how the transcription and translation process would produce these mechanisms which would the cause is equilibration happen again known as proteins.

(Refer Slide Time: 36:52)



So, let us look at this in a little more details. So, in a nutshell transcription where double stranded DNA you can see this DNA here getting converted into a single stranded DNA and the enzyme that... So, you use enzyme for this conversion and this is known as the RNA polymerase, so it synthesis is the messenger RNA as just in here. So, the RNA synthesis done by RNA polymerase molecule from DNA to RNA where a much more compressed go to is develop based on reading the DNA molecule and then you have this translation, which is done by this intelligent Nano machine we called ribosomes, which translates the sequence of the bases in the m-RNA to the corresponding proteins and these proteins then perform various functions inside and outside the particular cell. So, this is an essentially the amino acids sequence which you develop from this m-RNA here which has been developed further from the DNA.

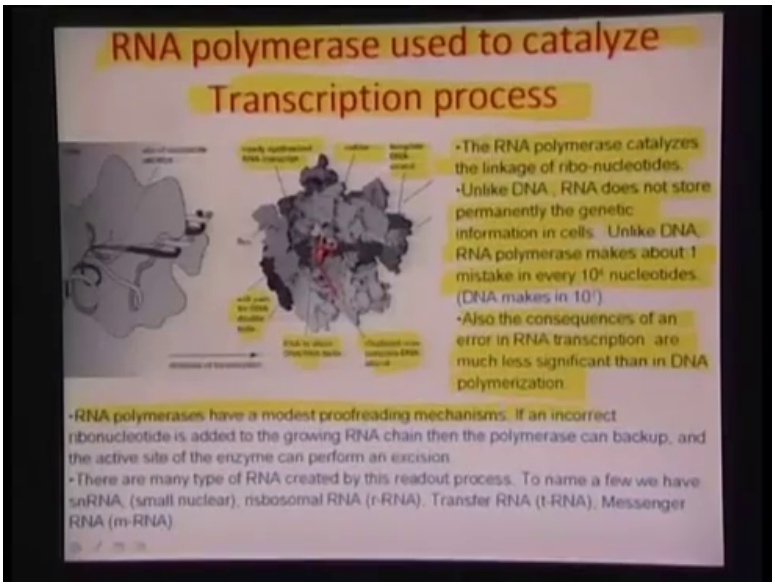
(Refer Slide Time: 37:56)



So, let us look on a step by step bases of how this whole process can be complex. So, this is essentially the transcription process as you seen here, let first talk about what exactly it is all about. So, this is double stranded DNA structure that you can look into this whole structure here is a double stranded DNA structure, this is the corresponding RNA structure that has been formulated, the only difference between a DNA and RNA is that RNA is a single strand and suppose to DNA and then the group thymine here on the DNA is kind of replace by a another group called uracil on the NA. So, the four bases in ribose sugar RNA or ACUG as supposed to ACTG in the DNA.

So, the process is essentially starting when the cell copies a particular portion of it is DNA nucleotide sequence into an RNA nucleotide sequence. I am going to look into the details of how this happens in the next slide the information on the RNA although copied into another chemical form is still in the same language and therefore, it is known as writing this strand script or transcription DNA. The name transcription is because the language is same essentially there is no difference in the language of the DNA or the RNA except that there are few groups and they here and there which are different and also the very fact that this RNA is a single standard structures, single standard molecule.

(Refer Slide Time: 39:18)



So, we use an enzyme here which would be able to read what is there on the DNA and produce a compressed information as in RNA. So, the enzyme is known as RNA polymerase and this is use catalyze the transcription process. So, essentially the RNA polymerase is nothing but, a set of proteins and it contains about some protein moieties and some RNA molecules in term. So, here itself this kind of gives the complicated structure of an RNA polymerase and you can see here in this particular region that the DNA change is kind of opening and coming out and then there is an alternate RNA chain, which is evolving from here.

So, there are several steps which are used by the RNA polymerase to carve messenger RNA from DNA and I am actually going to just look into the mechanical aspects of this process step by step and so what you need is that you need to bite into the DNA at a portion, where the signal which... So, there is a signal which comes into the cell, the signal maybe in terms of chemical group or moiety which can be able to identify a portion of the DNA from which you can replicate the corresponding to m-RNA.

So, that eventually after the translation process and the endoplasmic reticulum, the m-RNA, the sequences out the protein which is useful for equilibrating the state of the cell back to it is normal. So, there is a deviation in the environment of the cell due to which there is change in equilibrium of this particular cell due to some reasons and then there is a signal which is generated, because of this change in equilibrium which goes and bites on the portion of the DNA from which you are essentially trying to build up the messenger RNA and then once the messenger RNA build up it produces a protein which would go into that change the

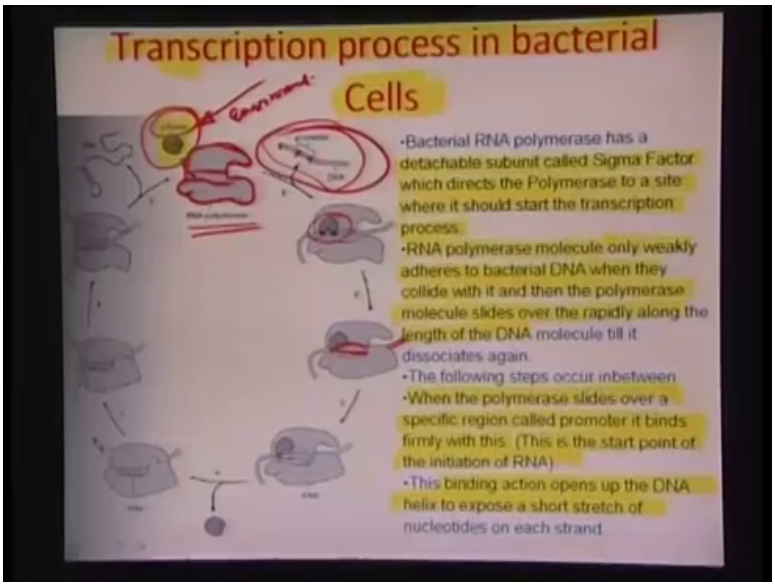
environment and try to bring it back to the equilibrium position or equilibrate the whole cell that is the whole fundamental of this transcription translation process.

Let us look at how? So, the RNA polymerase catalyzes the linkage of ribonucleotides and like DNA, RNA does not store permanently the genetic information in cells. So, some a characteristic the RNA polymerase has first have to be evaluate the for understanding how the RNA polymerase is would do this translation this transcription work. So, and unlike DNA RNA polymerase makes about one mistakes in every 10 to the power 4 nucleotides by doing this. DNA it is about 1 in 10 to the power 9. So, this one mistake in every 10 to the power 4 nucleotides would mean that there are several mistakes that the RNA can do while getting replicated from the DNA. And so it is not that vital or that important as the replication of DNA process.

Essentially a replication of a DNA would cause of mutation it is a permanent change resulting from this information change in the DNA. Whereas, there are thousands of RNA's which have produce any and any point of time from the DNA structure and therefore, it is not that critical to have a very high fidelity or very you know low error rate is essentially. So, you can manage within about one error in every 10 to the power 4 nucleotide easily and still nothing appreciable changes, essentially the protein coded from this m-RNA would have a slightly different sequence and then that really is not very critical, because there are several of this proteins which are being produce every second by the cell also the consequences of an error in RNA transcription of much less significant the in DNA polymerization as I just talked about.

So, RNA polymerase has a modest proof reading mechanism if an incorrect ribonucleotide is added to a growing RNA chain. So, you have a RNA chain which is growing on the bases of the template molecule of the DNA then the polymerase can back up and the active site of the enzyme can perform an excision. So, therefore, there is a cleavage of the inappropriately added nucleotide as and when the RNA polymerase is able to sense it there is a confirmatory change and based on that there are many type of RNA's which could be created ribosomal small nuclear transfer or messenger different forms RNA.

(Refer Slide Time: 44:26)



So, this here illustrate how this process happens. So, this structure here is essentially the RNA polymerase enzyme I actually call at magician hat. So, here there is this chemicals signal called and this is essentially a described for the bacteria cells. So, the transcription process here that I am going to describing just about few minutes, a few seconds in bacteria cells. So, there is a chemical signal which would come to this particular RNA polymerase enzyme, which would tell about where exactly on the DNA you should be able to bite. So, that you can start producing the messenger RNA.

So, this chemical signal is also a function of the environment. So, it is called sigma factor technically this is the name that is given to this chemical signal. So, this chemical signal comes and it kind of sits on this RNA polymerase and the RNA polymerase essentially goes into reading the DNA. So, this is DNA molecule here as you have seen the RNA polymerase kind of goes into reading the DNA through it is a sigma factor. So, we have to find out first the site from which the transcription process should start, the RNA polymerase should be able to start the transcription or it should be able to nowhere to start the coding into m-RNA from the DNA the transcription.

So, let suppose with the sigma factor the RNA polymerase kind of bites into this DNA is bites into this DNA region and then it loosely slights on this DNA from this end all the way up to an extend when it hits, so called promote a region is promote a region on the DNA is a kind of chemical orientation is structure, wherein the RNA polymerase and the sigma factor specially would be able to bite on the DNA molecular it is able to bite on the DNA and open the chain. So, that is called the promoter region of the DNA.

So, let suppose this is this DNA chain and this enzyme along with this sigma factor it is kind of sliding pass. So, the enzyme is the upper and lower half and you have the DNA chain which is going into this and coming out of other area and kind of slights closely over this until it hits of a promote region, where it kind of bites and holds back firmly. So, there it opens then chain of the DNA it is a double stranded DNA. So, it is opens that chain of the DNA.

So, therefore, if you really look at this process you have a detachable sub unit sigma factor it directs the polymerase to a site very start the transcription and the RNA polymerase is molecule only weekly at here's to the bacterial DNA when the collide with it and then the polymerase molecule slides over rapidly along the length of the DNA molecule until it hits the promoter region and there it kind of binds firmly or more firmly over the specific region, because of the sigma factor.

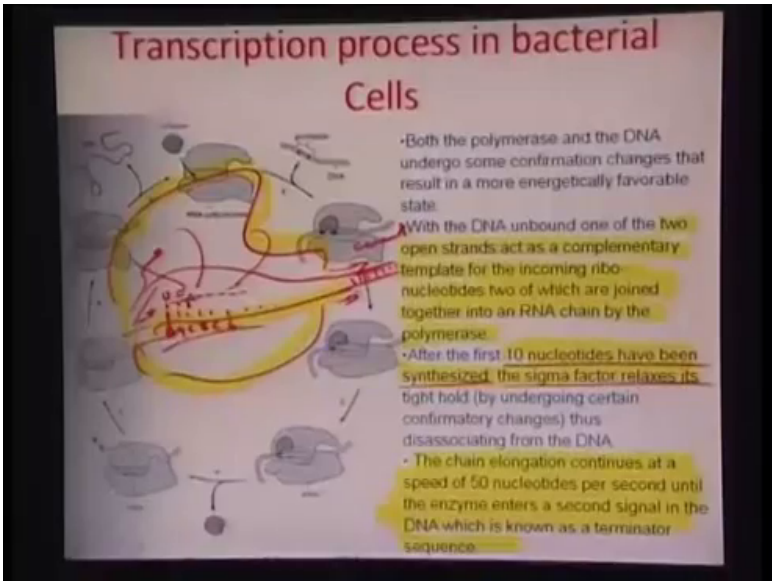
So, when the polymerase slides over the specific region called promoter it binds firmly with this and bind by binding we are meaning we are trying to cut the DNA chain open into two strands it is removing the hydrogen bonds between the base pairs on the DNA cutting into two different strands. So, this starts the initiation of the RNA transcription process or the RNA coding process.

So, once this is happen does you are able see here when the sigma factor has gone and found promoter region and it has been able to open this chain. So, the DNA is open chain here. So, when it comes out on the other side of the RNA it again kind of change back or zigzag, because thermodynamically that is the most favorable state once it goes outside the byte in mechanism of there in a polymerase. However in the region where the sigma factor has kind of bit and closely on the DNA there is a chain opening like this and these two are open from which one half of the chain would be used for templating the other half of the molecule. So, you are using this lower half let say for producing this is a template to sequence the m-RNA or the messenger RNA in the revere sequence of what is there in a DNA.

So, therefore, we can think of a compression of information this model is something where we can illustrate this or we can define this is a we do not need all the information that is there on the DNA, but we really need a selective specific region of the DNA which would be able to code a protein and that signal when it comes it kind of automatically compresses all this information’s into a really useful domain. So, it is a kind of signal compression on how many base pairs starting from what point ending to what point is really needs to be a transcripted as

a messenger RNA. So, the binding action opens up the DNA helix here and exposes the short stretch of nucleotides on each strand and so therefore, there is a templating mechanism which has taking place now.

(Refer Slide Time: 50:15)



So, whenever the two open strands acts complimentary templates for the incoming ribo-nucleotides two of which are joined together into an RNA chain by the polymerase enzyme. So, the polymerase essentially furnishes by assembling the ribo-nucleotides which are continuously coming in that... So, there are lot of ribo-nucleotides which are around in the solution in the nucleus. So, they continuously keep on diffusing through the RNA. So, as and when suppose there is a C group on the opened up DNA, it should corresponding to a G group on the nucleotide.

So, there is essentially a capture point for G to have come and suture there, then there is a let say a following group on the parental RNA which you have or when the parental DNA which are trying to copy, where they was C before and now there is a G following C. So, corresponding there would be a C nucleotide with sits next to the G nucleotide and these are individual nucleotides, the polymerase what it does is it sutures these to nucleotides and so therefore, it kind of starts zipping the DNA in the internal region where it has already opened up, but this zipping is action is by the individual nucleotides which are now trying to kind of get templated or arranged by this template DNA molecule.

Let say I will just illustrate this a little bit more clearly here, let say you have a DNA which is opened up, the chain has opened up like this and you have a templates sequence here, which

is corresponding to AGCT and then you have again CG something like that. So, there are lot of nucleotides which are rushing in here, so let say this side has been identified is an A definitely a nucleotide which would be independently let say you know uracil we know there are RNA there is a uracil.

So, therefore, there is no thymine in this region as an independent nucleotide it is more in the uracil form that would sit on the A then there would be a C or a cytosine which sits on the guanine here and then the RNA polymerase is essentially has a role in bonding these two therefore, there is a guanine which sits on the cytosine here, hydrogen bond on the RNA again suture this C to G. So, this RNA molecule is slowly develop on the template here, that you can see and this is what is happening inside the RNA polymerase is enzyme.

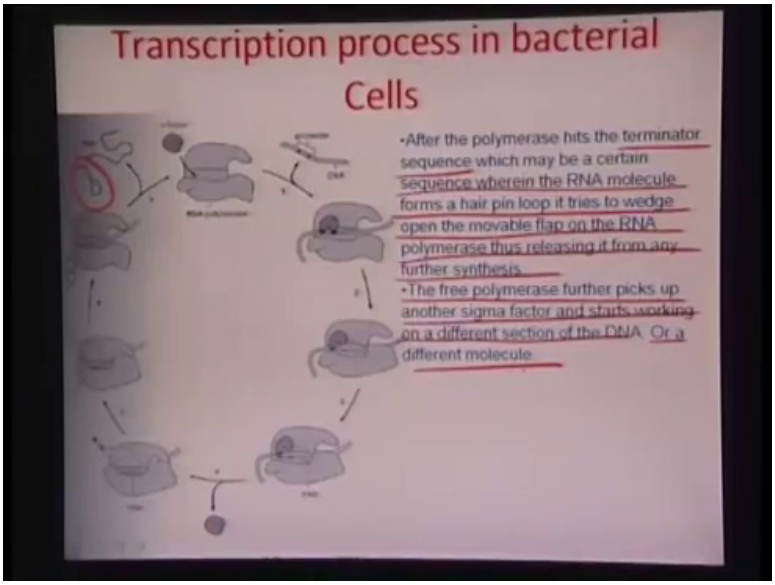
So, really after the first nucleotide first ten nucleotide sequences are kind of synthesize sigma factor relaxes itself by holding it less tightly now. So, it kind of on you know kind of grips are loosens out it is grip on the DNA. So, that is no more question of cutting open the DNA, you have already initiated as zipping up process and then it is thermodynamically favorable to kind of keep zipping up as the RNA rapidly, the RNA polymerase end of rapidly moves through the DNA chain. So, really after the first ten nucleotide sequences as you can see here in the sigma factor no longer place any role biting it is the RNA polymerase sliding in the auto opening of the chain which promotes this process.

And continuously the nucleotide sequence as the m-RNA is been generated from this template here that I showed just about few minutes back which has been opened up inside this whole RNA polymerase structure. Now so therefore, the chain elongation continues and there is of course, a rate aspect to this along this elongation, normally typically this suturing rate and you know the nucleotides carrying exampled over the template would take close to about 50 nucleotides per second or so. So, this is a finitely definable rate of production of the m-RNA. So, speed of production is above 50 unit's nucleotides per second.

So, now, once the processes started of course, you know m-RNA is define is a size define molecule there has to be some kind of a signal to the enzyme which let us it know that it has to stop and it has dissociate the m-RNA out and that is closing template as you have seen here the DNA template which have zipped up which are zip token in this region as you are feeding the template within the molecule was actually zipped back into zipped close as it goes cross the RNA and tries to emanate out of the RNA. So, this is the point from which it is emanating.

So, the zipping up again occurs when it emanates, the only open region is inside the RNA polymerase, let say this is the RNA polymerase molecule, the only open region is inside this RNA polymerase molecule. Now, you have to somehow let the RNA polymerase realize that the sequence that it has to make is complete and you should no longer produce any m-RNA. So, there are sequences on the m-RNA which are also known as stop codons which kind of the few sequence at least three or four different three different kind of sequences, where a combination of nucleotide is kind of hit upon, where there is a conformation change the enzyme the RNA polymerase as a hits upon this three base pair uniquely combine three base pair region call the stop codon region of the m-RNA. So, whenever that happens...

(Refer Slide Time: 56:04)



So, this is essentially also known as a terminator sequence, where the chain kind of terminates from getting manufacture. So, essentially this is a combination of three nucleotide on the m-RNA on the produce m-RNA on the template, where the RNA polymerase kind of gets a conformational changer or jerk because of which that is a see here of this templating activity, the RNA that is develop, the messenger RNA that is developed on the template kind of detaches from the template.

So, typically there could be mechanisms like let say the RNA molecule hits upon some kind of a hairpin loop look like structure here. So, the molecule that have just emerged because of this stop codon in that region it formulates some kind of a hair pin loop and so because this loop has been formulated which is not normal or not a normal sequence that can be a certain jerk or a conformation chain which can also cause this detaching movement to happen, whatever it is the changing conformation of the RNA polymerase is what we are vision to

have. So, that we can terminate the you know the chain fabrication process. So, that is exactly what happens here forms a hair pin loop and it tries to which open them only flap on the RNA and should releases the RNA from any further synthesis.

And so the free polymerase is welcome to pick up another sigma factor and start the DNA sequence signal accordingly on a different section of the DNA. So, the whole idea is that the proteins which are being produced or a function of what sigma factor or what chemical signal comes into to bite a certain region from which all it needs to hit upon is a terminated sequence for defining the length of the m-RNA that we are produce. So, it is very, very vital and interesting process of information compression that even the cells or the nucleus on the cell within intelligent molecule RNA polymerase can execute.

So, this kind of brings us to the end of this lecture I would like to illustrate this concept in a little more details and also start with another very important translation topic which would give us the conversion between the language of the DNA or the language of the RNA into the language of amino acids of proteins.

Thank you.