Bio – Microelectromechanical Systems

Prof. Shantanu Bhattacharya

Department of Mechanical Engineering

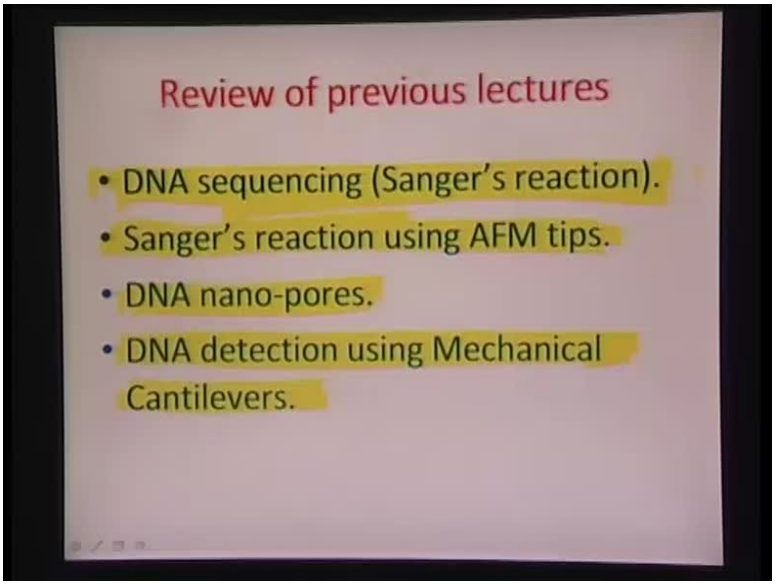
Indian Institute of technology, Kanpur

Module No. # 01

Lecture No. # 22

Hello and welcome back to this lecture 22 of Bio-Microelectromechanical Systems. Let us quickly review of what we did last time.

(Refer Slide Time: 00:22)



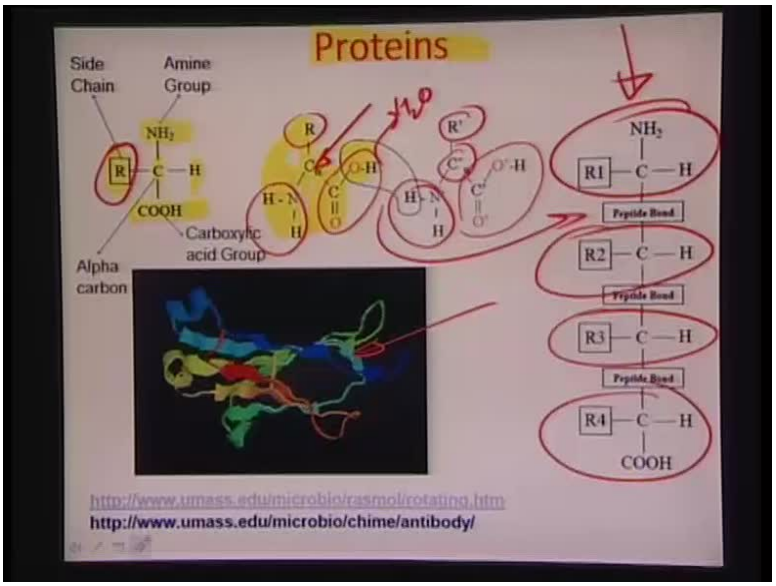
We talked about DNA sequencing using William Sanger's process, where we actually have a chain terminated ddNTP set of groups and four different reactions. As we know, the dd or the dideoxy NTP is essential for terminating the chain of the replication process. So, you can have various lengths terminated at 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 identify what really is the particular group at that instance or at that length from the overall readout.

So, essentially, you can actually sequence a DNA by looking along its chain, the ddNTP. Whether it is an ATP or CTP or GTP or thyminetriphosphate, it is able to cleave and terminate the chain replication reaction. We also had a little glimpse of the technologies, the most modern evolving technologies, where this Sanger's reaction can be used. The

transport of DNA could be done on a molecular level using an STM pen or an AFM tip. We further tried to explore DNA nano-pores, where you basically see, how in nature, self-assembled proteins over lipid bilayers can be used as DNA nano channels.

Similarly, you could also use silicon oxide on insulator platform to create a thin nano-pore over this platform and then use it successfully to translate DNA from a rich to a lean side using an electrophoresing potential. So, this can also be an identification protocol for sequencing DNA. As we talked in detail, we discussed some of the other alternate strategies of detection especially mechanical detection. Whenever you talk about mechanical cantilever, you essentially have diving board like structures. So, you basically have a change in surface energy causing a deflection delta Z, which is given by Stoney's equation. From there, you can actually find out what is the change in surface stress causing a certain delta Z value that you can detect; otherwise use optical or other means. Therefore, this summarizes what we did last time. I would like to now look at proteins. So, we will do some preliminary introduction of proteins today. We also like to correlate how this fantastic nano machinery of a living cell translates the information, which is available as a compressed DNA or RNA into an amino acid sequence or a protein.

(Refer Slide Time: 03:27)



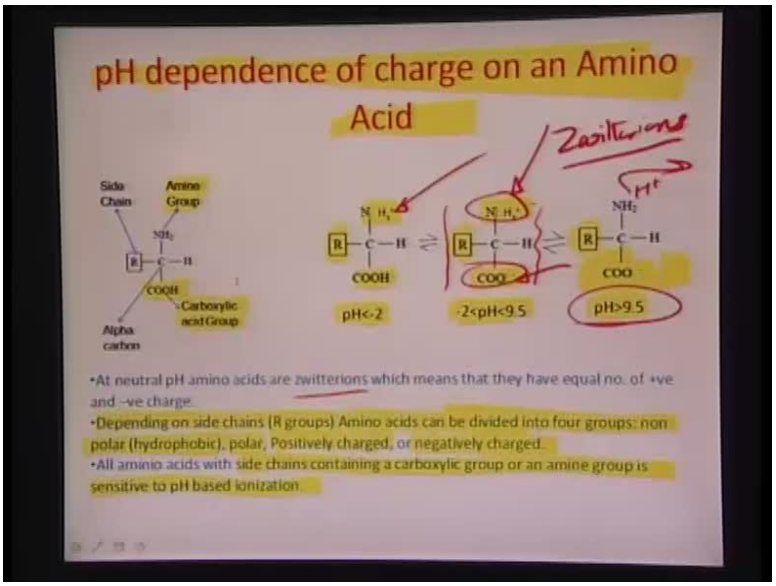
Let us look at what a protein really is. So, if you look at this figure right here, a protein really is a bunch of different amino acids. What is an amino acid? You have an alpha

carbon and we have a carboxylic group. Here, you have a side chain that can be a polar or a nonpolar moiety. You have an 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 an NH2 group on another side of an alpha carbon and that is the basic idea of an amino acid.

Now, the way these amino acids bond together is through these amide bonds. So, let us say, this (Refer Slide Time: 04:14) is one of the amino acids with the C alpha; the alpha carbon right here. This is the amine group and this is the carboxylic group. Essentially, the COOH is a carboxylic group on the acid and the other amine also has another group. Let us say, R dash on the side chain here, it is R and this alpha carbon is connected to an NH2 and also a COOH. So, there is a bondage between the OH on the COOH and NH2. Thus, producing a water molecule and you have, what you call a peptide bond. This essentially is something called a peptide bond, C single bond N and there is a double bond on one side with O N on the other inside. You have a single bond H. That is what essentially a peptide bond is.

Several such amino acids bond to each other with different end groups, may be with these peptide bonds to form arrays or chains of molecules called proteins. So, protein is basically defined as a chain of amino acid interconnected using this peptide linkage. So, this can be essentially a model of how a protein would typically look like, in a solution of a certain kind. So, one important point about protein is that it undergoes rapid conformational change because of R group. The R groups, as you see here can be polar or nonpolar depending on the type of applications or depending on the type of various amino acid groups that are present in this world. So, there are about twenty different such amino acid groups which are categorized into polar, nonpolar, ionic, non-ionic etc.

(Refer Slide Time: 06:07)



Let us look at some of the interesting things, the pH dependence of such an amino acid. If you really look at this carboxylic moiety here, it is a COOH and it is an acid group; carboxylic acid group and you have an amine group. So, protein when put in a certain pH can absorb an excess hydrogen ion onto its NH forming NH2 to form NH3 plus. It can depend on it; if the pH is slightly basic and generate hydrogen ion. Thus creating COO minus. What is interesting here is - it is a function of pH.

Suppose, the pH is very acidic it is the less than 2 of this hydrogen ion concentration. So, in this instance, as you see here, because of the excess available hydrogen ions in the solution, there is a tendency of the amine group NH2 to protonate to get into NH3 plus. It is one aspect of how positive charge can be produced on the amino acid. Now, there is also a range of pH. Let us say, in this particular case here (Refer Slide Time: 07:04), it is between minus 2 and 9.5. So, it is from slightly acidic to a reasonably high basicity. In this particular pH range, there is a tendency of the carboxylic group to protonate because there is an excess of OH minus concentration. Therefore, there is a tendency of the COOH to protonate and are converted into COO minus, as you can see here in this case.

Now, in this condition also (Refer Slide Time: 07:08), the total amount of positive and negative charges on this particular moiety is equal and opposite to each other. So, they are essentially known as zwitterions and that is a term to denote these kind of moieties, where there are ion centers within the same molecule; but then, the amount of positive

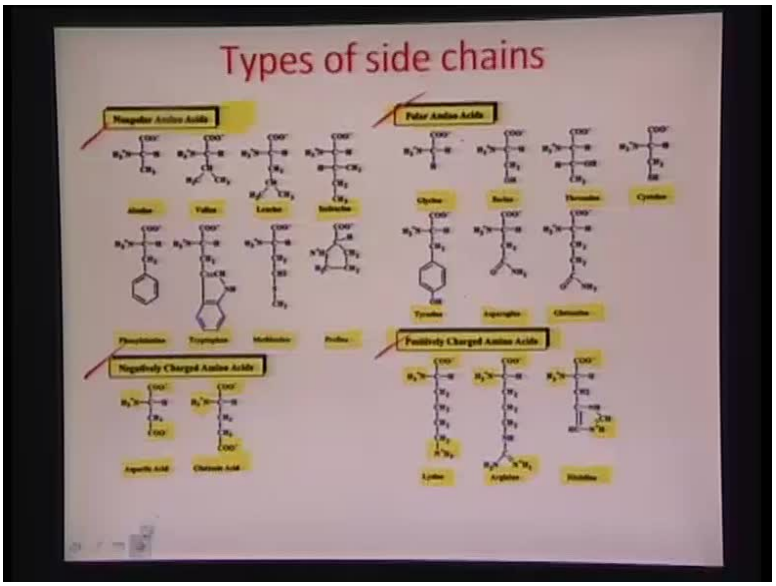
and negative charges are equal and opposite. So, they are called zwitterions and they have equal number of positive and negative charges.

Now, if you change the pH into more basic, the pH goes above 9.5. In that case, there is a tendency of the excess hydrogen ion, which was already formulated in the zwitterionic state to protonate into the solutions. So, the H plus goes out into the solution because the basicity is high enough for any positive ion in the positive charge hydrogen moiety on the amine group to present. So, it protonates and here again, as you see, the overall charge of the amino acid reduces to negative. So, the charge state of an amino acid really depends on the pH. If it is very highly acidic, you have a positive charge on the acid. If it is in a certain range, where they have equal amount of positive and negative charges, the amino acid essentially behaves as a zwitterion. If the pH is carried on, from there to more basic, there is a tendency of the protonation to happen again.

You essentially have a negative charge on the particular amino acid. So, depending on the side chains of this R group here; this is very critical, the R group may be polar, may be non-polar depending on whatever it is. Whatever its nature is, there is a tendency of the proteins to change their shapes and size by folding and essentially this is also known as a confirmation change. It is technically known by that name, but with the way that protein formulates into a close loop or a structure is really a function of what are there on the R groups on the protein. These side chains of the proteins are very critical to determine the interaction of the protein in an aqueous, a non-aqueous bounded environment.

So, amino acids depending on the side chain, be divided into four groups; non-polar or hydrophobic polar. The separated charges are positively charged, where there is an additional H plus ion on the particular amino acid. Negatively charged, where there is an additional minus charge on the carboxylic group by sending out an H plus into the solution. So, they are essentially based on the R groups. We can divide all the amino acids into 4 different groups.

(Refer Slide Time: 11:19)



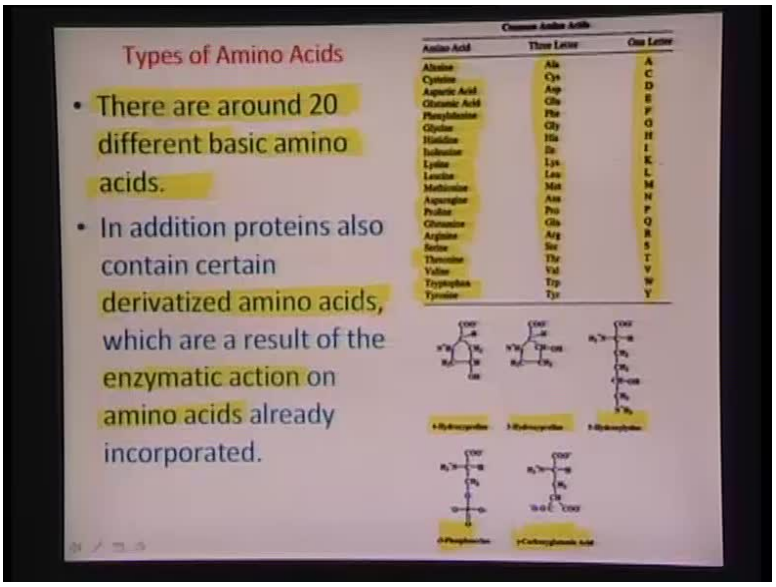
As we already learnt that these side chains containing carboxylic groups and amino acid groups, are really sensitive to pH based ionization and that is one aspect. So, let us look at what are the R groups. If you can categories all the different amino acids, they can essentially be drawn out as about twenty different amino acids. So, all the proteins that you see around are the thousands of chains or thousand types of chains that are formulated from these twenty basic amino acid groups that are present. Therefore, you can say that you can categorize the side chain of such an amino acid into a nonpolar moiety like Alanine, Valine, Lusine, Isolusine, Phenylalanine, Tryptophan, Methylamine, Proline. Essentially, these are the nonpolar amino acid; R groups on the amino acids.

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 electronegativity 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 R groups can be having different names like the glycine. I have said about Threonine, Cysteine, Tyrosine, Asparagine and Glutamine. These are some of the polar amino acid groups.

You can also divide the R groups into negatively charged amino acids and they can be Aspartic acid of Glutamic acid. You can see in these two cases and then you also have positive. So, in the negatively charged amino acid, as you see, there is an excess amount of negative charge centers over the positive charge center. Here also, it has an excess amount of negative charges in addition to the only positive charge that you have here.

So, these are essentially a function of pH. It states various pH states, but there is of course, an additional negative charge here. You have categories of positively charged amino acids like Lysine, Arginine and Histidine. You can see there is negative charge center, but then essentially, there are two amines, which at a certain pH would be able to… Even in its zwitterionic form, in the true essence or in the same pH range, it has two charge centers for positive charges and one for the negative charge So, these are categorized in the positively charged amino acids.

(Refer Slide Time: 14:11)



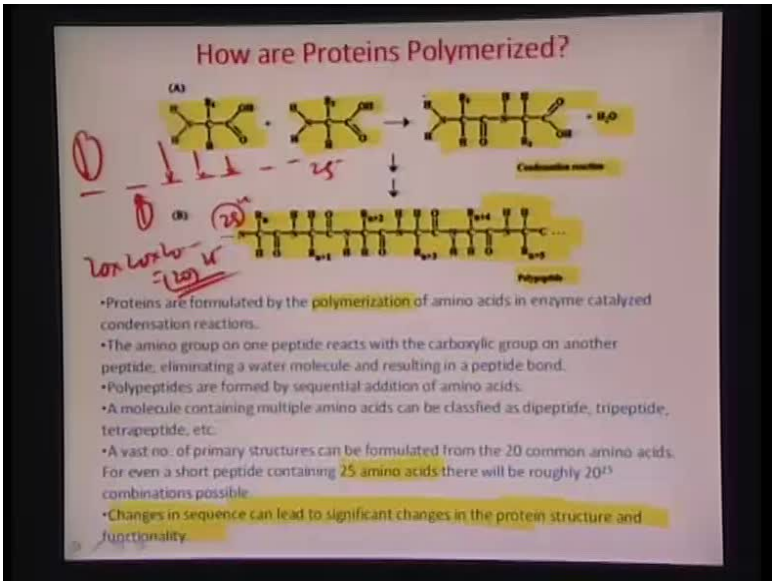
In a nutshell, depending on these type of side chains, twenty different amino acids can be categorized into various kinds like nonpolar, polar, negatively charged and positively charged. What is also important here is that let us just do some commentarial to find out the kind of sequences or the total combination of sequences that are available to us, which can define the whole protein world essentially.

There are around twenty different amino acids. So, this is kind of a list of all these twenty different amino acids. There are polar, nonpolar, negative or positive charges.

You just saw in the last slide about these and some of the synonyms, which are used either in three letters or one letter. As you can see here, the proteins also contain certain derivatized amino acids, which as a result of enzymatic actions on amino acids already incorporated. It is because of some enzymatic actions on these principal amino acids. There are certain derivatized forms in addition to the twenty basic amino acids. These are represented here, but I am not going to go into this. It is beyond the scope of this course to go into this enzyme derivatized amino acids, but there are lot of these enzyme derivatized in addition to the twenty basic amino acids that we have.

So, if you want to do a commentarial, let us say, we are talking about a 25 amino acid sequence. Each of these twenty-five places can have about twenty different amino acids. So, if we can repeat a sequence, then the total number of permutations is about 20 to the power 25 moieties. Such a vast amount of arrangements is available only for a 25 amino acid sequence.

(Refer Slide Time: 16:18)



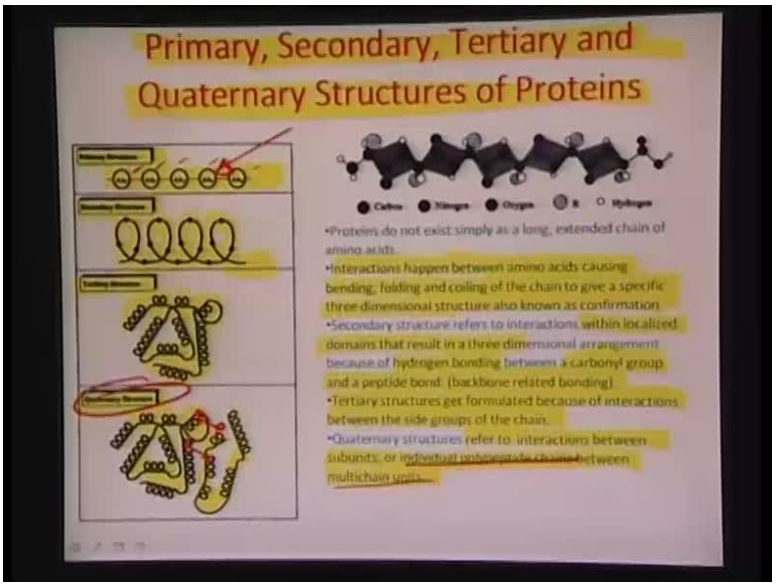
You have twenty groups like these, I am sorry; you can have twenty groups like these. You can have twenty amino acids in each of these areas. So, essentially 20 into 20 into 20 and so on till the 25th term. So, 20 to the power of 25 combinations and this is a vast amount. So, if you wanted to predict, what is one sequence out of these 20 to the power of 25 sequences of amino acids is really a huge number to deal with.

So, these are essentially the primary structures of the protein. 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 are in a sense a polymerization process. Here, two amino acid groups are condensed together in enzyme catalyzed condensation reaction. Why we call it condensation? Wherever there is a disbursement of H2O moiety or a water molecule, we call that reaction as condensation reaction. Here, this processes enzyme catalyzed. So, there is enzyme which will 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 acids from mRNA sequence in the translation process. They also use enzymes to cleave bonds or to make the peptide bonds. So, it is enzyme catalyzed condensation. We already know that a H2O moiety goes out on taking the OH from the carboxylic on one, H from the NH2 on another, but this process happens only in the presence of an enzyme and so a peptide bond is formulated.

Here, in this illustration, it pretty much shows how the systems are. So, you have two amino acid sequences here and then there is an enzyme, which would convert this. There is a condensation reaction. So, there is a H2O molecule, which is eliminated. You can have polypeptide chains like these, where you have several of these peptide linkages and more than one amino acid. So, proteins can be this large (Refer Slide Time: 18:39). Therefore, depending on how many peptides, the molecule is classified either into dipeptide or a tripeptide or tetrapeptide and so on. It also determines how many amino acids you have. So, 1 protein or 1 peptide linkage is associated with 2 amino acids. So, you could find out the total length of the protein chain based on categorizing into a di or tri or tetra peptide and so on.

We already know that how many commentorials we can process. It is possible only in a 25amino acid sequence of protein and this small sequence of protein can have about 20 to the power of 25 combinations. If you can club all these, possibilities of each place being filled by twenty different amino acids. So changes in sequence of a protein can lead to significant changes in the protein structure and the functionality. So, if the sequence is altered slightly, you can have a altogether different structure of the protein in interest.

(Refer Slide Time: 19:56)



What is also important here is that if we really categories proteins into different kind of structures, they can be classified principally into: the primary, secondary, tertiary and quaternary structures. Let us just investigate a little bit about why they are called so. What are the differences between these cartoons here? It illustrates very nicely about these four different structures.

So, proteins do not really exist simply as a long extended chain of amino acids. It is hardly possible to exist in a natural state as a long chain. One of the reasons, why that is so is that because there are a lot of OH groups and also there are lot of side chain interactions. There is also a tendency of hydrogen on the NH2 or COOH the OH on the COOH to interact with the solvent. So, there are additional hydrogen bonds, which are created in between the molecule and the solvent in question. So, confirmation really is a very dynamic process. The overall structure that happens is a very complicated structure because of the several hydrogen based interactions between the protein moieties and its surroundings or even molecules within the same protein moiety. Now, primary structure is very important in understanding, what is the sequence of the amino acids and proteins? So, essentially primary structure is the basic foundational structure. It gives a kind of relationship between what follows in terms of amino acid sequences. Although, it does not exist in nature, but it is always wonderful to represent a protein with its primary structure to find out the amino acid sequences that it has. So, interactions between the amino acid chains would cause the proteins to bend and fold. So, interactions happening

between the amino acids cause bending, folding and coiling of these chains. Eventually, what emerges is a specific three-dimensional structure and this is known as confirmation, as I have been discussing this before. So, secondary structure, as you can see here in this illustration is really formulated within localized domains because of the presence of hydrogen bonding and particularly because, there is a carbonyl group and a peptide bond. So, in between the protein and backbone itself, there is a lot of bonding. It is because of these two moieties. So, peptide bond on one side, then the carbonyl group on the other side. So, secondary structure is a coiling and it is like this (Refer Slide Time: 22:50). So, you can say that it is a very fine analogy and I keep giving that straight. Let us say, hosepipe was coiled in a manner and coiling is by means of either stressing in heated environment or something.

Now, whenever you stretch it totally in tension, put it in tension and stretch it, this is the primary structure that it would have been, but the moment you release the two ends, it is going to recoil back. It is because, the stresses would try to align it back into the equilibrium and that is the secondary structure. Let us say, this is a telephone cord we are talking about. Now, you have a case, where you have just released enough. So that the structure is coiled, then you drop the whole structure. So, what would happen is that this coiling would start to bend among itself. 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. They are essentially confined to the same molecule. There is only 1 single molecule and there are interactions between different domains of the molecule rather than just adjacent interactions as happening in the secondary structure. So, this is what tertiary structure of protein is and then of course, because of interactions between side groups of the chain in different regions of the same molecule.

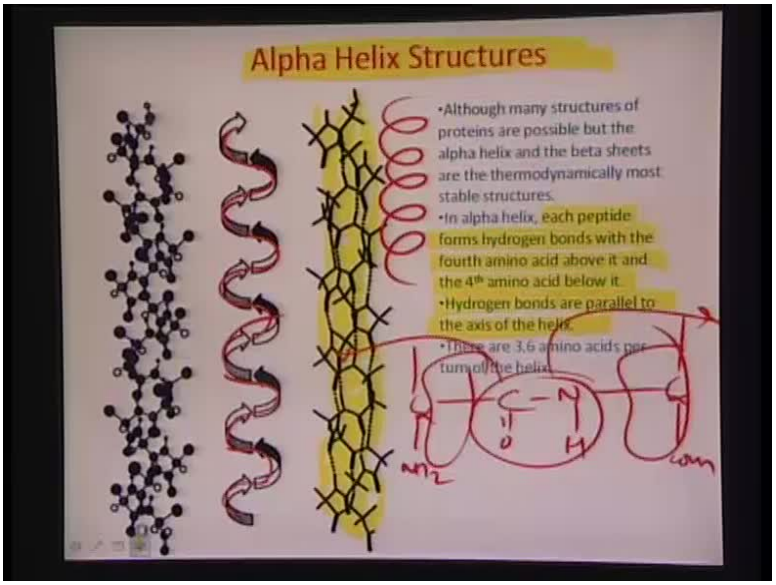
There is this quaternary structure, which is essentially referring to interactions between subunits or individual polypeptide chains and between multi chain units. So, for instance, there is one protein, which is present here in this area (Refer Slide Time: 24:50) and another protein is present independently in this area. So, there are two tertiary structures present close by. So, there is going to be interactions between the two structures. Therefore, between subunits of the same protein or individual polypeptide, chains are all laid side by side in a solution. That kind of multi chain interactions develop into what

you call as the quaternary structure. Therefore, starting from a simple to very complex proteins has increased degrees of complexities in their confirmations or in their shapes and sizes.

If you look at the proteins in general, there are many structures of these proteins, which are possible. We can principally classify the base units of these structures. Let us say, different unit cells in a certain lattice. We can also classify the proteins in terms of some base units, the repetition of which would essentially build the molecule as good in the crystal structure. As you know, the repetition of the unit 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. These are called analogs of unit cells and are known as the alpha helix and the beta pleats of proteins. So, you can either classify into an alpha helix structure… and I am just going to come to what an alpha helix really means. There is a beta plate, where you can say that the protein is kind of rotating and the molecule is like a big multiple S's compressed together. So, there are parallel sheets of amino acid chains, which are running at different levels.

(Refer Slide Time: 26:59)



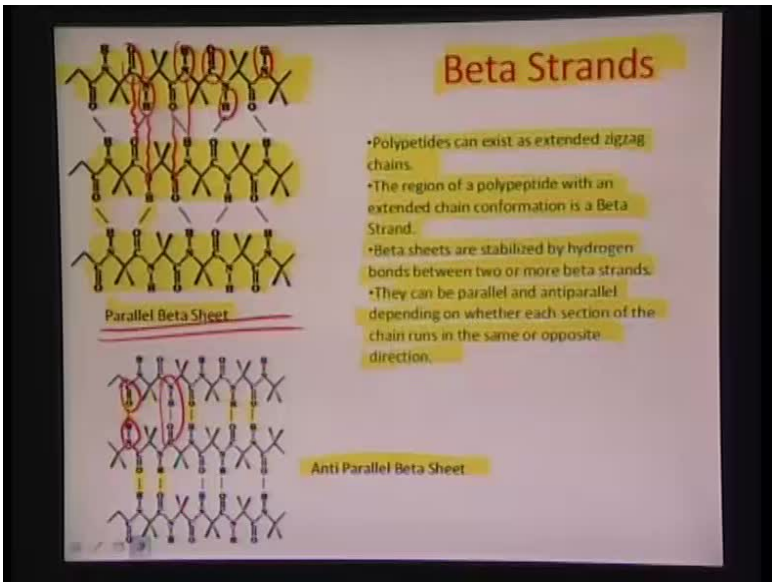
In Alpha Helix, as you see here, this is an alpha helix. Each peptide here forms hydrogen bonds with the 4th amino acid above it and the 4th amino acid below it. So, the way it goes is that this peptide linkage, which is essentially this CN double bond O and H. Then

you have the other portion of the protein in both sides here, 1 alpha carbon and things attached to it. So, here you have NH2 attached and here you have a COOH attached something like that. So, this peptide bond with the 4th amino acid in the sequence ahead and the 4th amino acid in the sequence at the back formulates hydrogen bonds because of which, there is a tendency of this molecule to form alpha structures. This is on a three dimensional basis that it is formulating the alpha. So, it is more like this (Refer Slide Time: 28:07) and that is the alpha structure that we are talking about. So, essentially, it coils just as any telephone cord would do. There are bondages formed between the peptide and every 4th amino acid on either end. So, a 4th amino acid in the bottom and 4th amino acid in the top would have hydrogen bonds essentially in 1 helical unit or in 1 unit of the helix that is formulated.

There are about close to 3.6 amino acids. In this kind of an orientation, hydrogen bonds would define the alpha helix in a manner, where one helix has containment of 3.6 amino acids. So, I mean there is no physical interpretation of 0.6. Only that 60 percent of 1 full amino acid is within the chain in additional to the three units and that is how you define the helix length of this particular structure.

One interesting factor here is that the hydrogen bonds are kind of parallel to the axes of the helix. That is that is the reason why this alpha kind of structure is formulated this elliptical. I would say this helical kind of structure is formulated because of all the bonds- essentially in all the different areas of the proteins- are parallel to the central the axis of the helix, which confines the molecule within this helical domain and that is what alpha structure is.

(Refer Slide Time: 29:56)

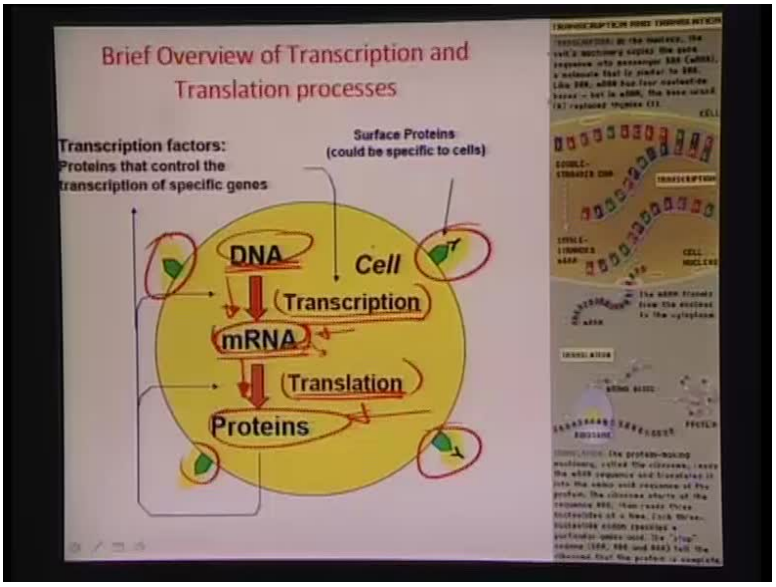


The second is the analog for unit cell. You may call the second structure that is most commonly used as a unit structure in protein of the beta strands. The way these beta strands structures can go is that they are like parallel sheets or they are like antiparallel sheets. There are two different kinds of beta structures, where these proteins would be laid on the top of one another like sheets of different amino acid sequences. As you can see here, these are parallel because in both the molecules here, the groups are together. This (Refer Slide Time: 30:32) is the C double bond O, this is the NH, this again is a C double bond O, this again the NH. So, you have to parallelality in the way these groups are placed on a single sheet of these beta sheeted structure. So that is why they are called as parallel beta sheets.

You may have an antiparallel configuration, where the C double bond on one would comprise or would face the NH on another or vice versa. So, this is an antiparallel kind of arrangement. Therefore, polypeptides can extend as zigzag chains and as you are seeing here, they are called as beta strands. The region of the polypeptide with an extended chain, the confirmation is known as a beta strand. Beta sheets are stabilized by hydrogen bonds. As you see here, there is a hydrogen bond between the C double bond O on the carboxyl and the NH on the amine group. Similarly, there is another one here between the C double bond O and the lower sheet, the NH on the upper sheet and so on. So, there are lot of hydrogen bonds in parallel or in anti parallel. Of course, the anti parallel has a lot more stability than the parallel. Therefore, they can be parallel or anti

parallel depending on whether each section of the chain runs in the same or the opposite direction. The beta sheets are stabilized by hydrogen bonds between two or more beta strands. This (Refer Slide Time: 32:03) is the antiparallel beta sheet and this is the parallel beta sheet. In a nutshell, these are the two structures: alpha helix and the beta sheets.

(Refer Slide Time: 32:05)



Now, after going into this little bit and 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? Today, it is probably one of the most fascinating ballgames of the chemists as well as molecular biologists because this in a sense gives the cell a physiological stability. So, it is like a genetic code, it is like a small computer cell, which can convert a compressed code of these ATCG sequences into a sequence of amino acids or proteins.

The proteins, as we know are more or less present in the cell membrane and which is also responsible for the state of health for the physiological state of the cell. The way it grows, replicates and everything is related to the surface proteins that come in surface of the cell. The process, which allows these proteins to be generated is situated deep within the nucleus of a cell, where the messenger RNA is essentially formulated from a section of DNA section on that small region in the chromatin. In the nucleus, there is a compressed sequence of helical coils of DNA, which is also known as chromosomes.

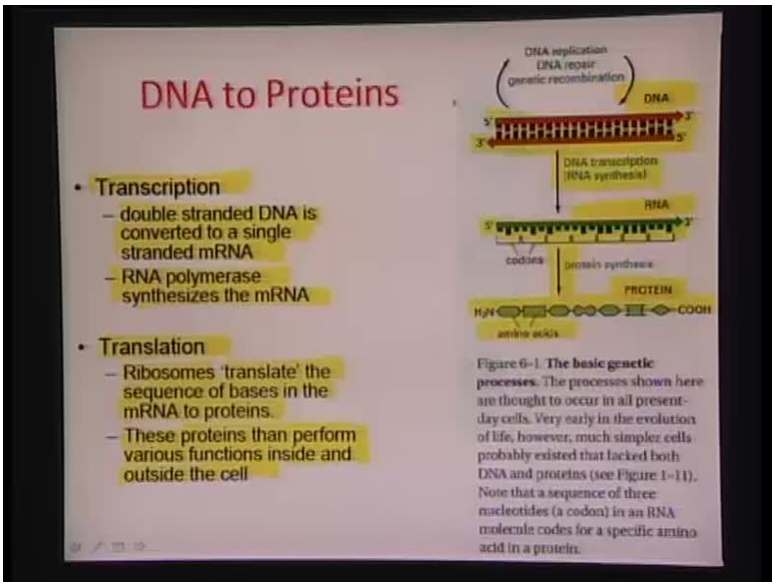
So, there is a process within the cell, where the DNA molecule gets converted into messenger RNA molecule. This is the compression algorithm and so the amount of data you have on the DNA is certainly much bigger in comparison to the amount of data you have on the mRNA. It is a much shorter information sequence than on the DNA. So, you have a compression algorithm here. So, you have messenger RNA and this process is known as transcription. So, you are making a transcript and you are not changing the language. The language is the same as ATCG and only in this case, the thymine is replaced by a neurocell. This is normally the case with all RNA moieties and of course, this is single stranded structure, which is formulated from the DNA and it is a double stranded structure.

The mRNA is generated within nucleus and then it goes all the way up to outside the nucleus into a place called endoplasmic reticulum. It is an organelle, which is there inside the cell and is essentially called the protein warehouse. So, there are intelligent nano machines called ribosomes on the surface of this organelle level of the cell, which essentially would take up a mRNA sequence and then convert that into a protein sequence. The way it does is a very fascinating. There are some amino style transfer RNAs, which are moving around in the solution. There is a group to group matching or a mismatching, which leads to the amino acid sequence as the mRNA or the messenger RNA sequences read at the nucleus level. Thus, this process of translation formulates amino acid sequence of the protein. Why it is called translation? It is because the language of the mRNA, which was more in terms of ACG and U is now converting into altogether different language. So, you are translating the information on mRNA into a sequence of amino acids.

So, it is a language conversion that is happening. So that would result in development of these surface proteins here. This is a function of how the cell would stay in a certain environment. So, the way that a cell reacts by having a signal from its environment would in turn expressed by or would try to be back into… So, the cell will be in equilibrium and it is based on whatever responses or stimuli it is getting from the environment. The way that the equilibrium happens is through producing a production of a set of proteins, which would exactly cause the deviation from the equilibrium to be balanced back. So that the cell comes back to the equilibrium and this depends on what is the existing physiological state, what is the deviation from that state and what would be

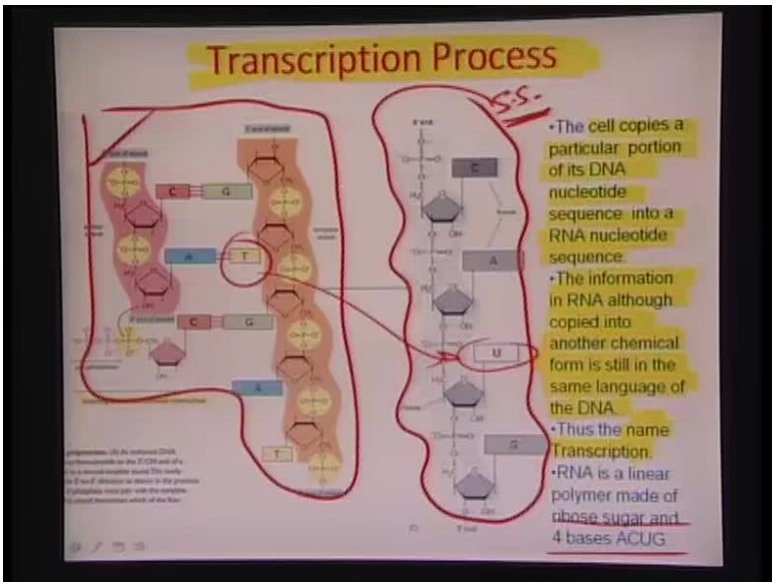
needed to control that deviation from that. These are three steps, which defines how the transcription and translation process would produce these mechanisms, which would cause this equilibrium condition to happen again, known as proteins. So, let us look at this in a little more details.

(Refer Slide Time: 36:57)



In a nutshell, transcription is where a double stranded DNA is getting converted into a single stranded DNA. We use an enzyme for this conversion and this is known as the RNA polymerase. So, it synthesis the messenger RNA here and so the RNA synthesis done by RNA polymerase molecule from DNA to RNA, where a much more compressed code is developed based on reading the DNA molecule. Then you have these translations, which is done by these intelligent nano machines. We call ribosomes, which translates the sequence of the bases in the mRNA to the corresponding proteins. These proteins perform various functions inside and outside the particular cell. So, this is essentially the amino acid sequence, which you develop on this mRNA here and which has been developed further from the DNA.

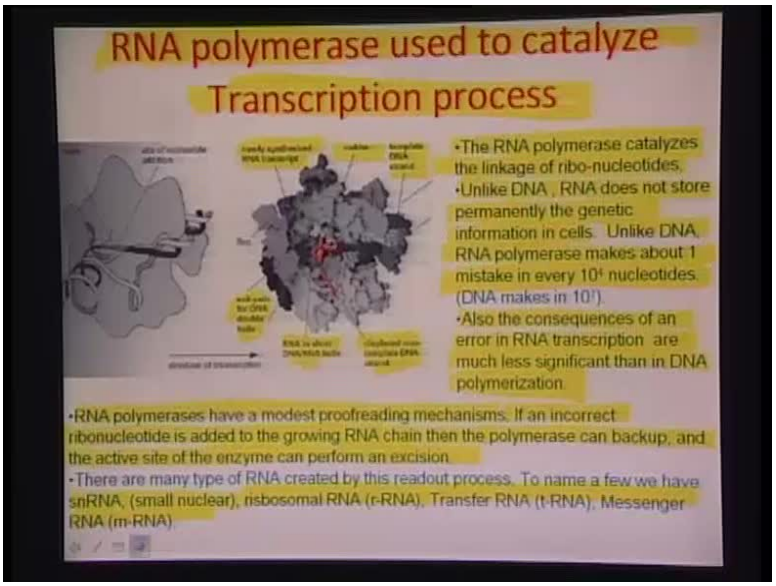
(Refer Slide Time: 38:01)



Let us look on a step-by-step basis of how this whole process can be accomplished. So, this is essentially the transcription process. As you see here, let us first talk about what exactly it is all about. So, this is double stranded DNA structure that you can look. 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 as opposed to DNA and then the group thymine here on the DNA is kind of replaced by another group called neurocell on the RNA. So, the 4 bases in ribose sugar RNA are ACUG as opposed to ACTG in the DNA.

So, the processes is essentially started, when the cell copies a particular portion of its 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 is although copied into another chemical form; it is still in the same language. Therefore, it is known as writing this transcript or transcription DNA. The name transcription is because, the language is same and essentially, there is no difference in the language of DNA or the RNA except that there were few group, which are different and also the very fact that this RNA is a single-stranded structural single strand molecule.

(Refer Slide Time: 39:24)



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

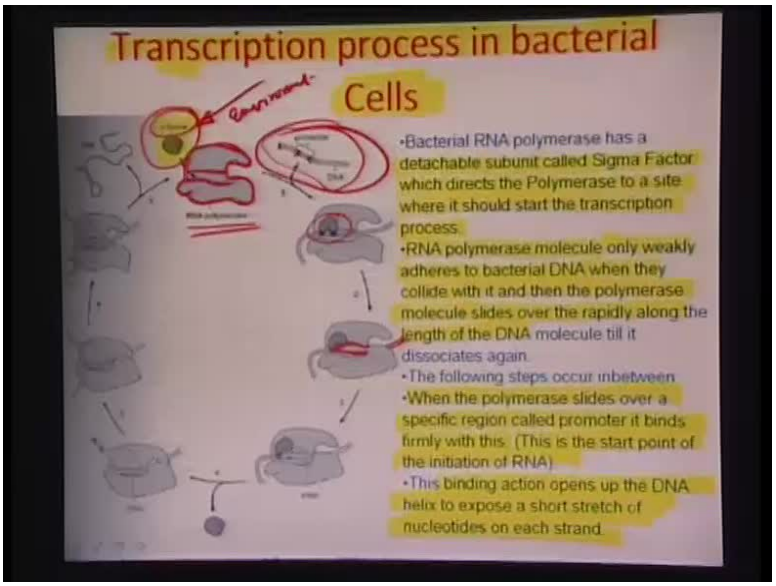
So, there are several steps which are used by the RNA polymerase to carve a messenger RNA from DNA. I am actually going to just look into the mechanical aspects of this process step-by-step. So, what you need is that you need to bite into the DNA at a portion. So, there is a signal, which comes into the cell. A signal may be in terms of a chemical group or a moiety, which is able to identify a portion of the DNA from which you can replicate the corresponding mRNA. So, eventually after the translation process on the endoplasmic reticulum, the mRNA sequences out a protein, which is useful for equilibrating the state of the cell back to its normal. So, there is a deviation in the environment of the cell and due to which there is a change in equilibrium of this particular cell. Due to some reasons, there is a signal, which is generated because of this change in equilibrium. It goes and bites on the portion of the DNA and from which, you are essentially trying to build up the messenger RNA. Once the messenger is built up, it

produces a protein, which would go into that changed environment and try to bring it back to the equilibrium position or equilibrate the whole cell and that is the whole fundamental of this transcription, translation process.

Let us look at how the RNA polymerase catalyzes the linkage of ribonucleic nucleotides like DNA. RNA does not permanently store the genetic information in cells. Some characteristics are that the RNA polymerase first has to be evaluated for understanding how the RNA polymer does this translation, transcription work. So, RNA polymerase makes about 1 mistakes in every 10 to the power 4 nucleotides and while doing this in DNA, it is about 1 in about 10 to the power 9. So, this 1 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. So, it is not vital or that important as the replication of a DNA process. Essentially, replication of a DNA would cause a mutation. It is a permanent change resulting from this information change on the DNA, whereas there are thousands of RNAs, which are produced at any point of time from the DNA structure. Therefore, it is not that critical to have a very high fidelity or very low error rate.

Essentially, you can manage about 1 error in every 10 to the power 4 nucleotides easily and still nothing changes appreciably. Essentially, the protein coded from this mRNA would have a slight different sequence. Then that is really not very critical because there are several of these proteins, which are being produced every second by the cell. The consequences of an error in RNA transcription are much less significant than in DNA polymerization, as I just talked about. RNA polymerase has a modest proof reading mechanism, if an incorrect ribonucleotide is added to a growing RNA chain. So, you have an RNA chain, which is growing on the basis of the template molecule of DNA. Then the polymerase can back up and the active side of the enzyme can perform an excision. Therefore, there is a cleavage of the inappropriately added nucleotide as and when the RNA polymerase is able to sense a confirmatory change. Based on that there are many type of RNAs, which could be create ribosomal small nuclear transfer or messenger different forms of RNA.

(Refer Slide Time: 44:32)



This illustrates how this process happens. So, this structure (Refer Slide Time: 44:37) here is essentially the RNA polymerase enzyme. I actually call it a magician's hat. So, there is this chemical signal called… This is essentially described for the bacterial cells. So, this transcription process here that I am going to describe in just about a few seconds in bacterial 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 and this chemical signal is also a function of the environment. So, it is called sigma factor and technically, this is the name that is given to this chemical signal. So, this chemical signal comes and sits on this RNA polymerase.

The RNA polymerase essentially goes into reading the DNA. So, this is a DNA molecule here and as you see, the RNA polymerases goes into reading the DNA through its sigma factor. So, we have to first find out the side from which the transcription process should start. The RNA polymerase should be able to start the transcription. It should be able to know where to start the coding into mRNA from the DNA transcription. So, suppose, sigma factor, the RNA polymerase bites into this DNA. This bites into this DNA region and then it loosely slides on this DNA from this end all the way up to an extent, when it hits the promoter region. This promoter region on the DNA is a kind of chemical orientation structure, where the RNA polymerase and the sigma factor specially would be able to bite on the DNA molecule. It is able to bite on the DNA and open the chain.

So that is called the promoter region of the DNA and let us suppose this DNA chain. This enzyme along with this sigma factor is a kind of sliding past. So, the enzyme is in upper and lower half. You have the DNA chain, which is going in and coming out of the other area. It kind of slides closely over this, until it hits up a promoter region, where it kind of bites and holds back firmly, so that it opens up the chain of the DNA. It is a double stranded DNA and so it opens up the chain of the DNA.

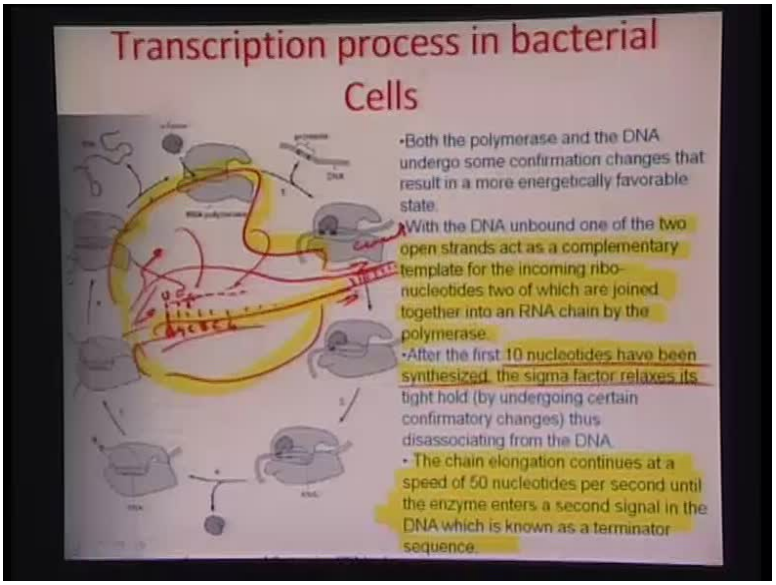
Therefore, if you really look at this process, you have a detachable subunit sigma factor. It directs the polymerase to a sight, where it should start the transcription. The RNA polymerase molecule weakly adheres to the bacterial DNA, when they collide with it. The polymerase molecule slides over rapidly along the length of the DNA molecule, until it hits the promoter region. It binds firmly or more firmly over the specific region because of the sigma factor. So, when the polymerase slides over the specific region called the promoter, it binds firmly with this and by binding 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 and cutting into two different strands.

So, this starts the initiation of the RNA transcription process or the RNA coding process. Once this has happened, you are able to see, when the sigma factor has gone and found the promoter region. It has been able to open this chain and so the DNA is open chain here. When it comes out on the other side of the RNA, it again chains back or zips up because thermodynamically that is the most favorable state. Once it goes outside the biting mechanism of the RNA polymerase, however in the region where the sigma factors bite closely on the DNA. There is a chain opening like this and these two are open from which one half of this chain would be used for tinplating the other half of the molecules. So, you are using this lower half. Let us say, for producing this as a template to sequence the mRNA or the messenger RNA in the reverse sequence like that of a DNA.

Therefore, we can think of a compression of information. This model is where we can illustrate this or we can define this as 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. That signal when it comes, automatically compresses all this informations into a really useful domain. So, it is a kind of single compression on how many base pairs starting from what point and ending to what point really needs to be a

transcripted as a messenger RNA. Alright, the binding action opens up the DNA helix here and exposes the short stretch of nucleotides on each strand. Therefore, there is a templating mechanism, which has started taking place now.

(Refer Slide Time: 50:21)



Whenever the two open strands act as complimentary templates for the incoming ribonucleotides, two of them are joined together into an RNA chain by the polymerase enzymes. So, the polymerase essentially furnishes by assembling the ribonucleotides, which are continuously coming in that. So, there are lot of ribonucleotides, which are going around in the solution in the nucleus. So, they continuously keep on defusing through the RNA.

Suppose, there is a C group on the opened up DNA, it should correspond to a G group on the nucleotide. So, there is essentially a capture point for G to have come and sutured there. Let us say, a following group on the parental RNA. You are on the parental DNA and you are trying to copy, where there was C before, now there is a G following C. So, correspondingly, there would be a C nucleotide, which sits next to the G nucleotide. These are individual nucleotides and the polymerase, what it does is - it sutures these two nucleotides and therefore, it kind of starts zipping the DNA in the internal region, where it has already opened up. This zipping action is by the individual nucleotides, which are now trying to get templated or arranged by this template DNA molecule. Let us say, I will just illustrate this a little bit more clearly here.

Let us say, you have a DNA, which has opened up like this (Refer Slide Time: 51:54). You have a template 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 us say, this site has been identified as an A; definitely a nucleotide, which would be independent. Let us say, we know in RNA, there is a neurocell. Therefore, there is no thymine in this region. As an independent nucleotide, it is more in the neurocell form that would sit on the A. Then, there would be C or a cytosine, which sits on the guanine here. The RNA polymerase essentially has a role in bonding these two. Therefore, there is guanine, which sits on the cytosine. Here, hydrogen bond and the RNA again sutures C to G. So, this RNA molecule is slowly developed on the template here that you can see and this is what is happening inside the RNA polymerase enzyme.

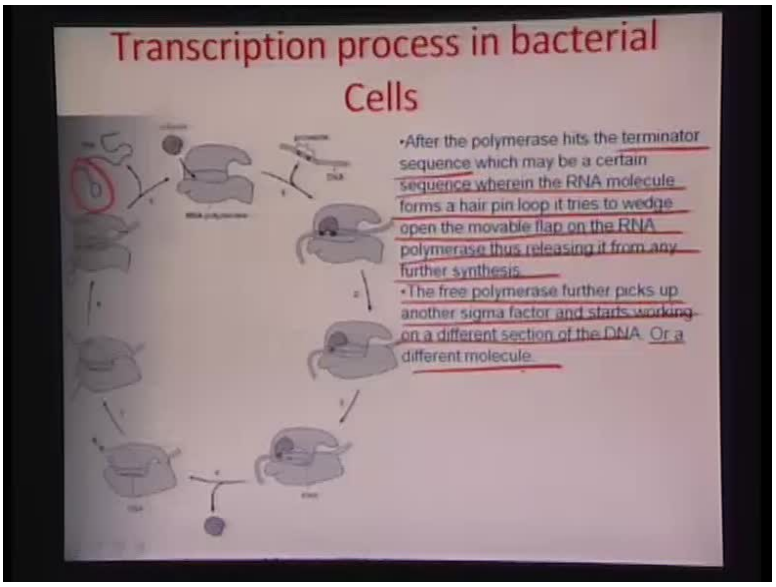
So, after the first 10 nucleotide sequences synthesized, the sigma factor relaxes itself by holding it less tightly. So, it ungrips or loosens out its grip on the DNA. So, there is no more question of cutting open the DNA. You have already initiated a zipping up process and then it is thermodynamically favorable to keep zipping up, as the RNA polymerase rapidly moves through the DNA chain. So, after the first ten nucleotide sequences, sigma factor no longer plays a role of biting. It is the RNA polymerase sliding and the auto opening of the chain, which promotes this process. Continuously, the nucleotide sequence, mRNA is generated from this template here that I showed just about few minutes back, which have been opened up inside this whole RNA polymerase structure. Therefore, the chain elongation continues and of course, a rate aspect to this elongation, normally this suturing rate and you know that the nucleotides is assembled over the template would take close to about 50 nucleotides per second. This is a finitely definable rate of production of the mRNA. So, speed of production is about 50 nucleotides per second.

Now, once the process has started, mRNA is a size-defined molecule. There has to be some signal to the enzyme, which makes it know that it has to stop, it has to dissociate the mRNA out and thus closing the template. As you are seeing here, the DNA template, which had zipped up open in this region. As you are feeding, the template within the molecule was actually zipped back or zipped close, as it goes across 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. Only open region is inside the RNA polymerase. Let

us say, this is the RNA polymerase molecule and 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; it should no longer produce any mRNA. So, there are sequences on the mRNA, which are also known as stop codons. There are few sequences at least three or four different kinds of sequences, where a combination of nucleotide is hit upon, where there is a confirmation change of the enzyme the RNA polymerase and as it hits upon this three base pair. Uniquely, combined three base pair region is called the stop codon region of the mRNA.

(Refer Slide Time: 56:10)



So, this is essentially known as a terminator sequence, where the chain terminates from being manufactured. So, essentially this is a combination of three nucleotides on the produced mRNA on the template. The RNA polymerase gets a conformational change or a jerk because of which, there is a seizure of this templating activity. The messenger RNA is developed on the template and it detaches from the template.

So, typically there could be mechanisms like the RNA molecule hits upon some kind of a hair pin loop like structure. So, the molecule has just emerged because of this stop codon in that region. It formulates some kind of a hair pin loop and this loop has been formulated, which is not normal or not a normal sequence. There can be a sudden jerk or a confirmation change, which can also cause this detaching movement to happen.

Whatever it is, change in confirmation of the RNA polymerase is what we are wishing to have. So that we can terminate the chain fabrication process and that is exactly what happens here. It forms a hairpin loop and it tries to wedge open the movable flap on the RNA. So, it releases the RNA from any further synthesis. So, the free polymerase is welcome to pick up another sigma factor and start the DNA sequencing accordingly on a different section of the DNA. So, the whole idea is that the proteins, which are being produced, are a function of sigma factor. What chemical signal comes into bite a certain region and from which all it needs to hit upon is the terminator sequence for defining the length of the mRNA that we are producing. So, it is a very vital and interesting process of information compression that even the cells or the nucleus in the cell with its intelligent molecule RNA polymerase can execute.

So, this brings us to the end of this lecture. I would like to illustrate this concept in a little more detail and 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 or proteins. Thank you.