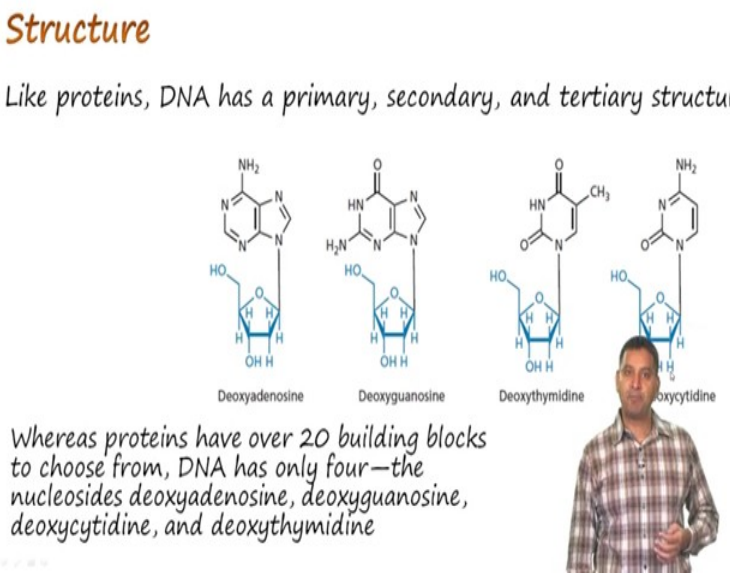
Medicinal Chemistry
Professor Dr. Harinath Chakrapani
Department of Chemistry
Indian Institute of Science Education and Research, Pune
Nucleic Acids – DNA Structure and Function

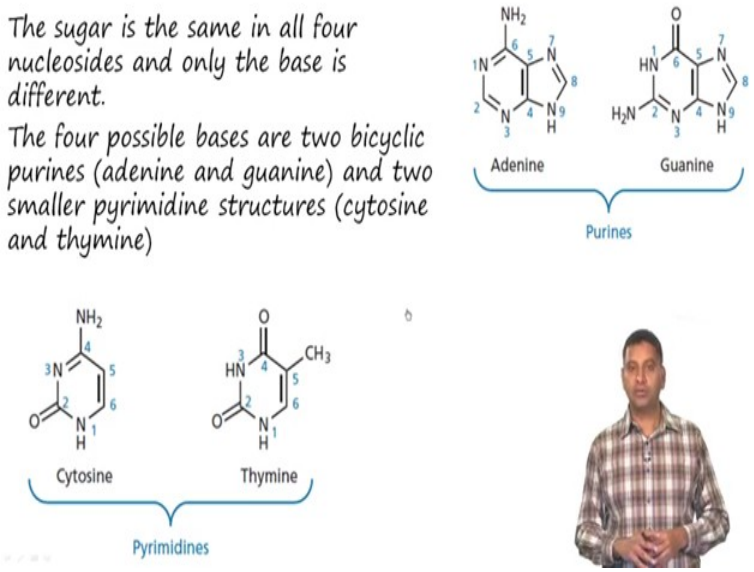
In today's lecture we are going to look at nucleic acids which are very important components of the functioning of a cell. So primarily today we look at how DNA functions and we will try to understand the structure of DNA. So as you may, as we have discussed previously the DNA carries the genetic information of a cell and what happens is that this genetic information needs to be transferred to RNA which is then subsequently transferred to proteins. And proteins are the workhouses of the cell. So in order to understand the entire functioning we need to learn a little bit about the basics of nucleic acids.

(Refer Slide Time: 0:59)



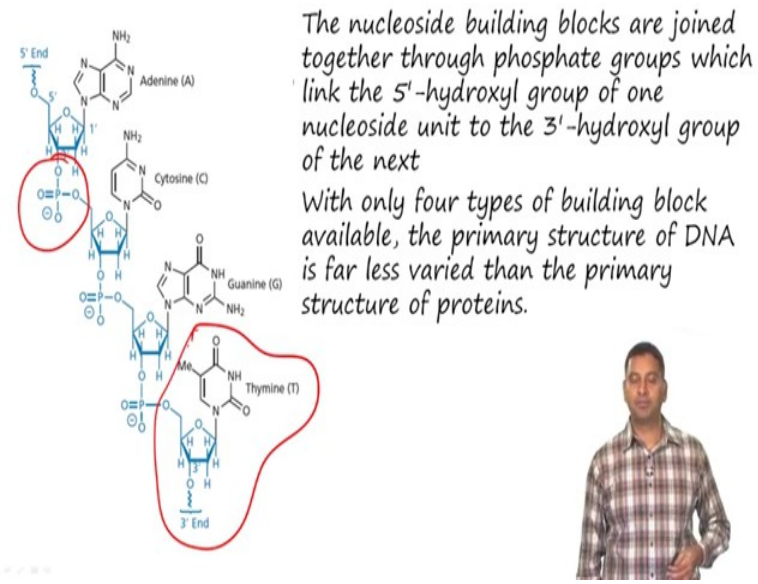
So much like the proteins which we have looked at previously, DNA also has primary, secondary and tertiary structures. So while proteins have about 20 building blocks which are 20 amino acids, DNA has only 4. That is, deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine, the structure of which are shown on the slide.

(Refer Slide Time: 1:28)



So if you look at the fundamental component of the DNA, DNA contains a base which are basically there are two bicyclic purines which are adenine and guanine, and there are monocyclic pyrimidine structures which are cytosine and thymine. The sugar is the same in all four nucleosides.

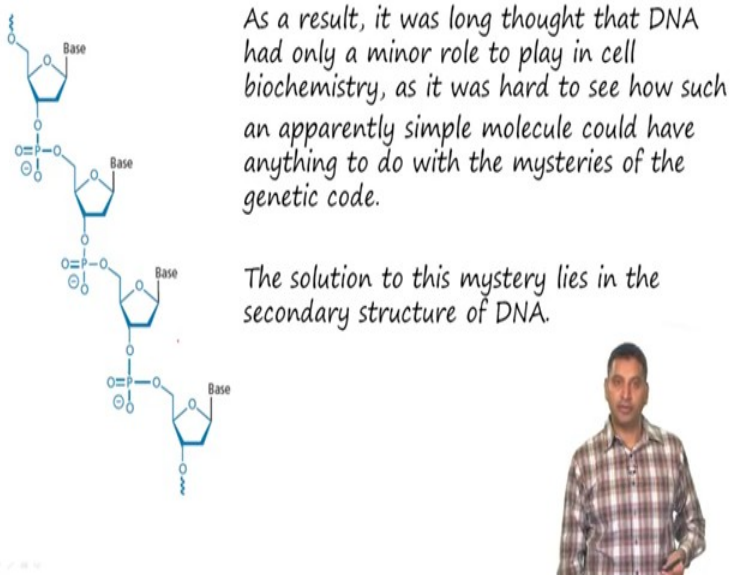
(Refer Slide Time: 1:51)



Now once this nucleoside building block is ready, the way it is joined together is through what is known as phosphate groups. So here is the phosphate group, here is a structure of nucleoside. For

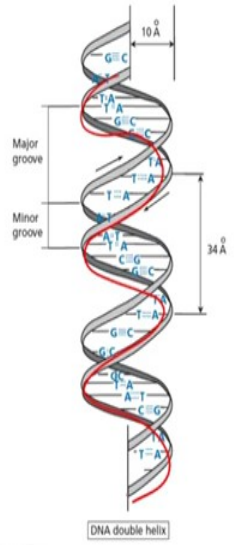
example, the analogue of thymine. And the way it is linked is through a phosphate diester. So with four types of building blocks the primary structure of DNA is far less varied compared to the primary structure of a protein.

(Refer Slide Time: 2:26)



So as a result it was, for long it was thought that DNA would have perhaps only a minor role to play in cell biochemistry because being such a simple molecule with only minimum amount of variation how is it that you can achieve complexity which is present inside the cell. This mystery is solved if one looks at the secondary structure of DNA.

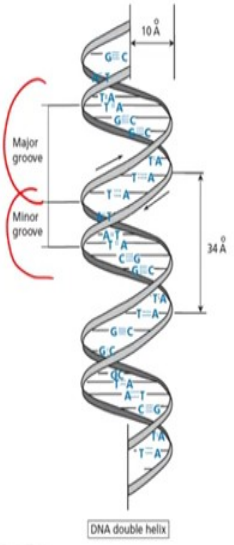
(Refer Slide Time: 2:50)



- Watson and Crick solved the secondary structure of DNA by building a model that fitted all the known experimental results.
- The structure consists of two DNA chains arranged together in a double helix of constant diameter

So the secondary structure of DNA was solved by Watson and Crick which is basically a double helix. I think all of you are familiar with this concept, so I will spend very limited time on this concept. But DNA exists as two single strands which are bound together as a double helix. So you have here the first strand, let's say it runs around here. And here is the, the other one is the second strand.

(Refer Slide Time: 3:23)

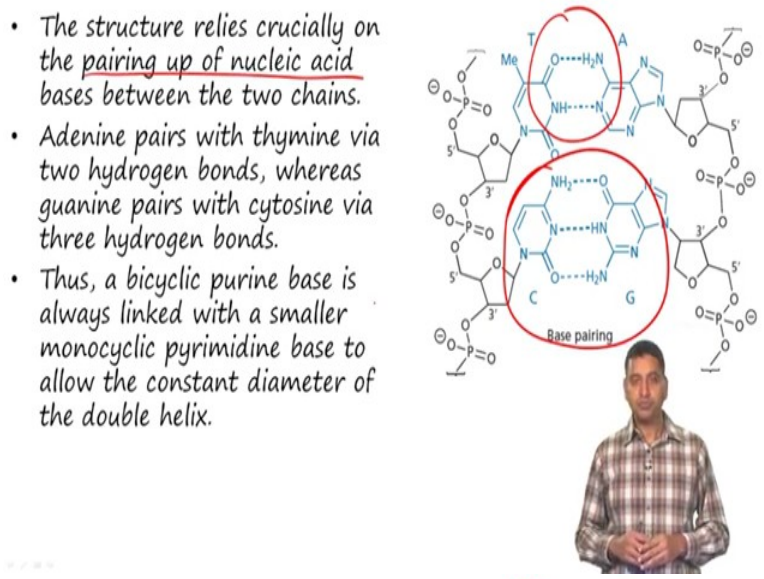


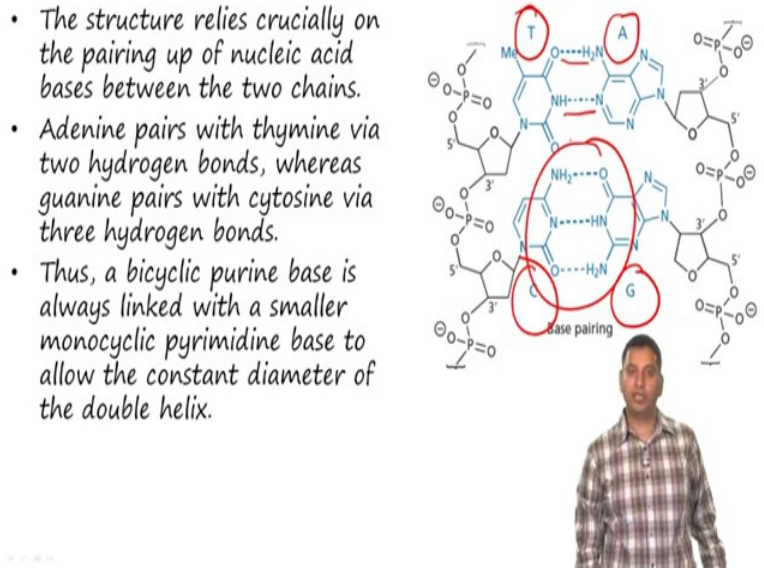
- The double helix has a major groove and a minor groove, which are of some importance to the action of several anticancer agents acting as intercalators

And what happens is that you have double helix has a major groove and a minor groove. And these are important when, in terms of, when we are looking at certain anti-cancer drugs which are known as intercalators. We will discuss this in detail later on in the course. But what these do

is that they can go and bind to the major groove or the minor groove and prevent or destabilize the DNA, and prevent the cancer cell from replicating.

(Refer Slide Time: 3:57)

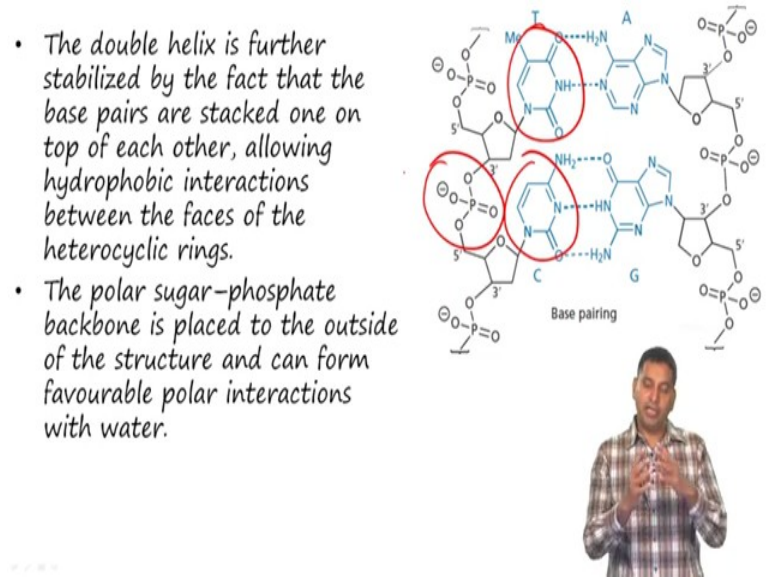




So also the DNA has, the interaction in DNA is by, through hydrogen bonds. The most important characteristic of DNA is that there is pairing of nucleic acid bases. Now what is being observed is that adenine always base pairs with thymine and guanine always base pairs with cytosine. And the nature of interaction between these two is through hydrogen bonding. Now between adenine and thymine there are two hydrogen bonds that occur. Whereas between cytosine and guanine there are three hydrogen bonds that occur. And this allows for a constant diameter of the double

helix. So you can imagine that this bicyclic purine base, once it base pairs you will have a constant diameter of double helix.

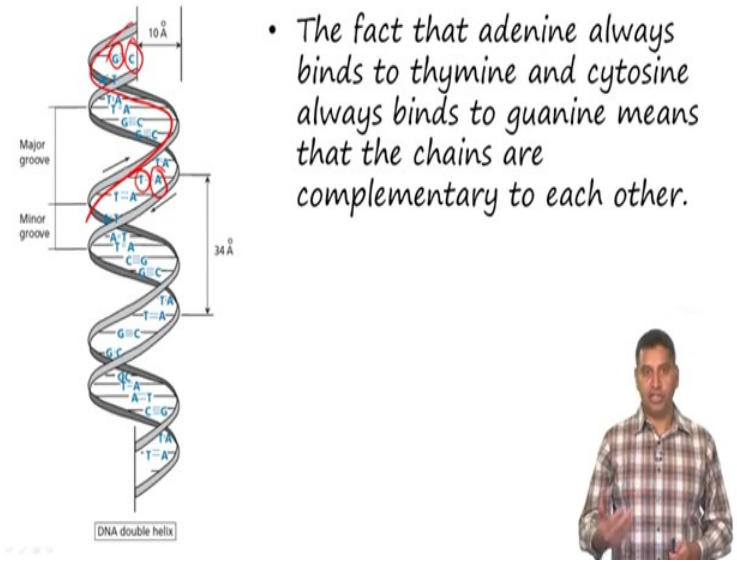
(Refer Slide Time: 4:50)



The double helix is also further stabilized by the fact that if you imagine that there is one base which is base pairing to the other one like this, and next one goes this way, and the next one goes that way, what happens is that these base pairs can actually have pi stacking interactions. So here is the base here and here is the base here. So you can imagine that they can be hydrophobic interactions which can occur, which also further stabilizes the double helix.

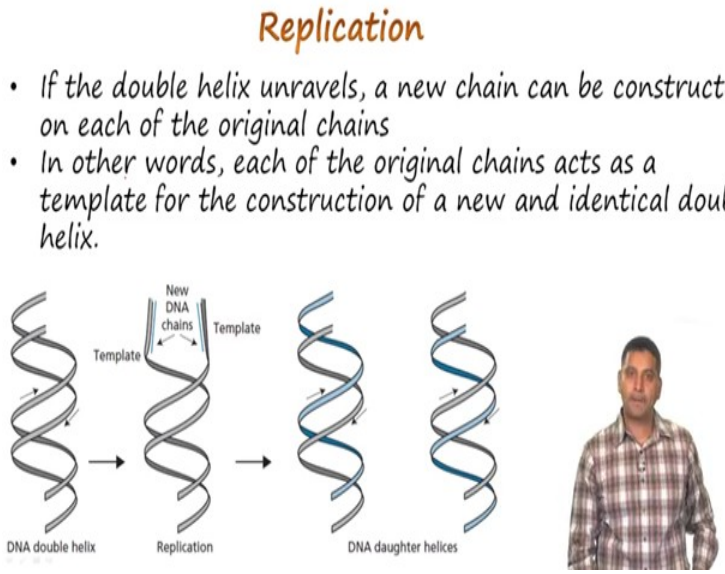
Lastly as we mentioned earlier there are these phosphate groups which link the DNA and these phosphate groups are actually highly polar in nature. So when DNA exists in -aqueous environment which has lot of water, this polar group helps in interacting with water and again further stabilizes the structure. So in summary the DNA double helix is a very highly optimized structure wherein there is hydrogen bonding, there is pi stacking or hydrophobic interactions and also polar interactions with water by the phosphate groups. All these together help in making DNA an extremely stable structure.

(Refer Slide Time: 6:04)



An important component of the structure of DNA is that there is complementarity. The means adenine always binds to thymine and cytosine always binds to guanine, which means that if you take any one chain and look at the sequence of one chain, the other chain will always be complimentary to the first chain. That means when you have guanine in the first chain, you will always have cytosine in the other chain. And when you have adenine in the first chain, you will always have thymine in the other chain. This is very useful because you have two strands which are complimentary to each other.

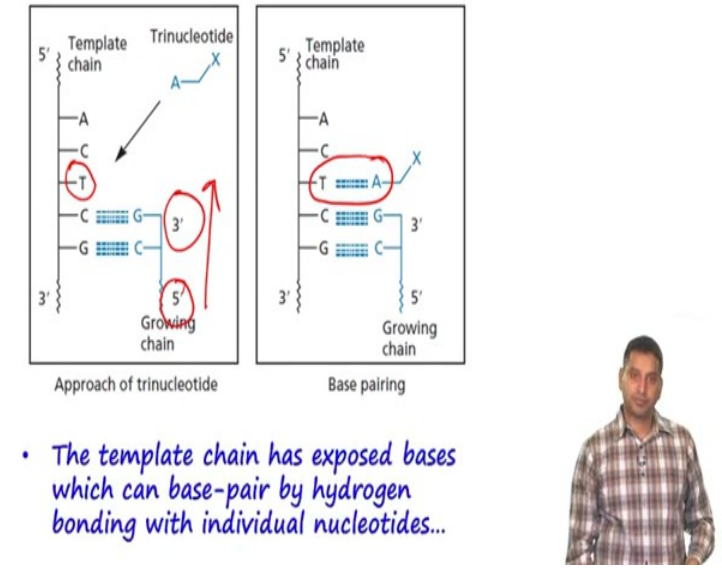
(Refer Slide Time: 6:51)



And when you have, when the cell replicates or when the cell divides, the first step is that the double helix unravels. And now when the new chain is constructed, the new chain uses the original or the parent chain as the template. And so when the DNA grows, it grows and it forms the complimentary base pair and once it replicates you have two DNA daughter helixes which are identical to the parent.

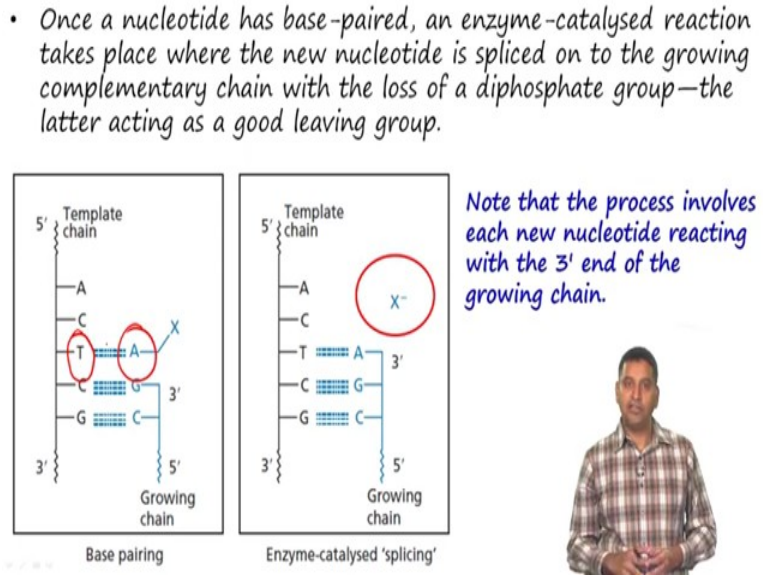
In other words, each of the original chain acts as a template for construction of a new and identical double helix. So what this ensures is that during cell division the genetic information that is present in the parent is actually transferred to the daughters.

(Refer Slide Time: 7:43)



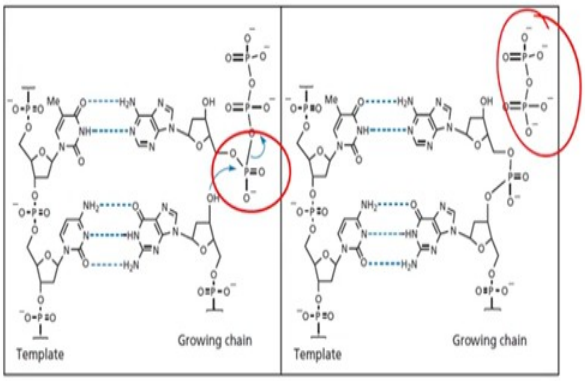
How does this process happen? What happens is that when you have a growing chain, you have a 5 prime end and a 3 prime end here is the growing chain, the incoming trinucleotide recognizes the base over here and as we mentioned earlier thymine is going to preferentially bind to adenine. And so therefore this interaction is highly favorable and so you have hydrogen bonding over here.

(Refer Slide Time: 8:19)



And therefore once the base is set in the right position, a reaction can take place which can kick out the leaving group, in this case diphosphate and form a new phosphodiester bond. So to summarize what happens is that in the growing chain, the next base that comes in, is always complimentary to the template chain. So the thymine always reacts with adenine and forms an intra-molecular hydrogen bond which then helps in making the chain elongation with a sequence specific manner. This is known as enzyme-catalyzed splicing.

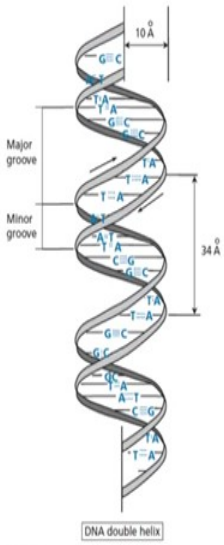
(Refer Slide Time: 8:58)



The chemical reaction that occurs is a displacement of a phosphate...

So to see the organic chemistry over here, what happens is that you have a phosphodiester bond which then, the alcohol in the sugar attacks the phosphodiester and kicks out a bisphosphate and produces bisphosphate here.

(Refer Slide Time: 9:16)



- How can only four nucleotides code for over 20 amino acids?

But you may wonder that since DNA has only 4 nucleotides how does it transfer the information to make 20 amino acids. So there is some mystery over here, Isn't it?

(Refer Slide Time: 9:32)

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCG	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

- The answer lies in the triplet code .
- An amino acid is coded not by one nucleotide, but by a set of three.

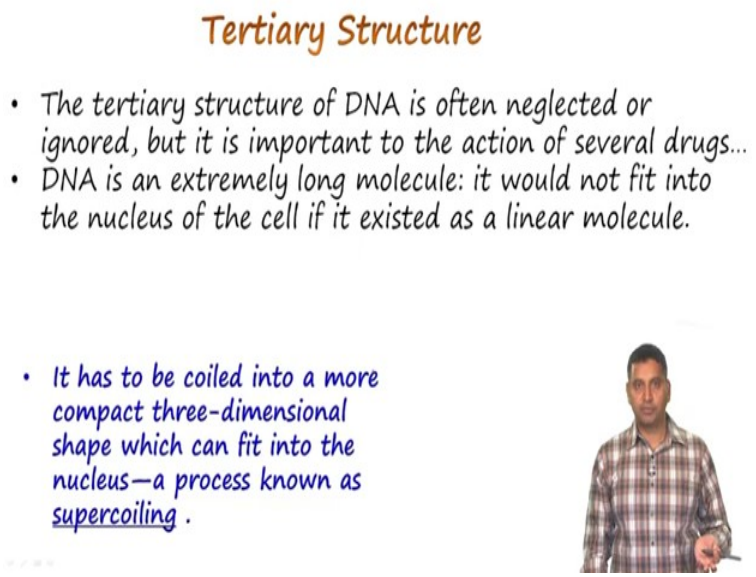
So the way this happens is that the DNA, single base pair of DNA does not code for a protein, but instead it is known, it has been found that there are three bases which code for an amino acid. So once you have these three bases such as UUU or CAC or GGC and so on, each of these codes are going to code for a single amino acid, one or more amino acids.

(Refer Slide Time: 10:06)

- There are 64 (4^3) ways in which four nucleotides can be arranged in sets of three—more than enough for the task required.

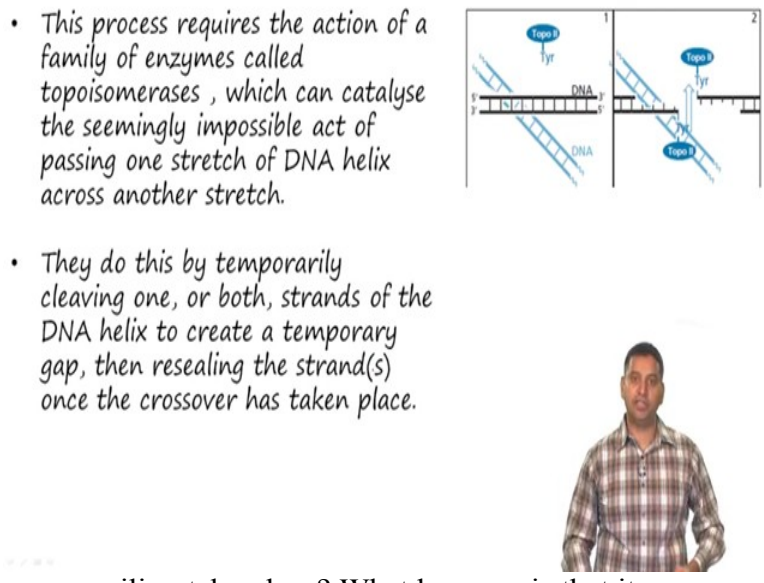
Now if you look at the combinations, there are four base pairs and so, and there are three base pairs which are coding for a particular amino acid. So if you look at the combination, it would be 64 ways in which four nucleotides can be arranged in sets of three. Now, since there are only 20 amino acids, this is more than enough for the task required.

(Refer Slide Time: 10:28)



Now let us look at the tertiary structure of DNA. So if one were to take DNA and extend the length of the strand, it is impossible for us to fit this in a cell. So therefore what happens is that it sort of coils into a more compact 3-dimensional shape which can fit into the nucleus. This process is known as supercoiling. And of course, needless to say the tertiary structure is often neglected or ignored, but it is important to the action of several drugs.

(Refer Slide Time: 11:06)



Now how does this supercoiling take place? What happens is that it uses an enzyme known as, a family of enzymes called as topoisomerases and these can catalyze the passing, the act of passing

one stretch of DNA double helix across another. The mechanism by which this happens is that they cleave both, one or both of the strands and create a temporary gap and then once the crossover has taken place, they reseal the strands.

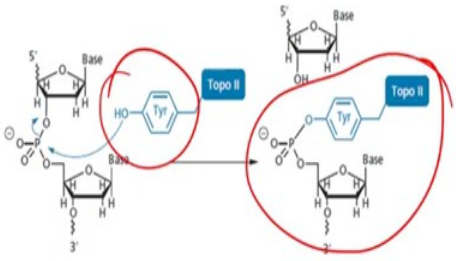
(Refer Slide Time: 11:39)

- Supercoiling allows the efficient storage of DNA, but the DNA has to be uncoiled again if replication and transcription are to take place.

What happens is that supercoiling allows for efficient storage of DNA, but keep in mind that the DNA has to be uncoiled again if replication and transcription are to take place.

(Refer Slide Time: 11:52)

- Topoisomerase II is a mammalian enzyme that is crucial to the effective replication of DNA.
- The enzyme binds to parts of DNA where two regions of the double helix are in near proximity
- Th e enzyme binds to one of these DNA double helices and tyrosine residues are used to nick both strands of the DNA

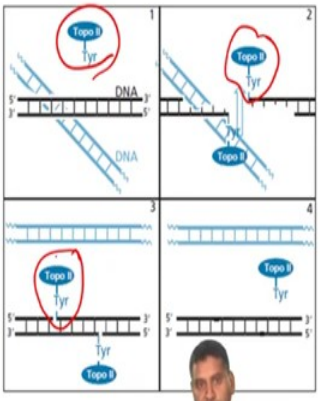


The way in which this topoisomerase works is it has a tyrosine which has a free alcohol or a phenol which then attacks phosphodiester bond and kicks out the alcohol. And what it does is it

forms a temporary covalent adduct with DNA which then subsequently reacts with proximal DNA and forms the other bond.

(Refer Slide Time: 12:18)

- A temporary covalent bond between the enzyme and the resulting 5' end of each strand results, thus stabilizing the DNA.
- The strands are now pulled in opposite directions to form a gap through which the intact DNA region can be passed.
- The enzyme then reseals the strands and departs.



What happens is a temporary covalent bond between enzyme and the 5 prime end of each strand is formed. And once this is formed, the strands are now pulled in opposite directions so that there can be a gap through which the other DNA can be passed. So this is a very systematic way of supercoiling. And this, after this is done, the nicks are then resealed by the enzyme.

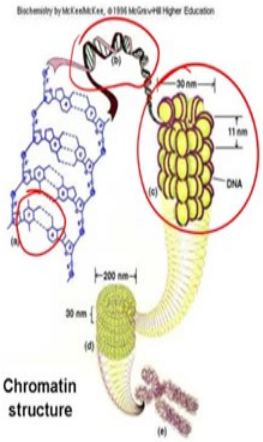
(Refer Slide Time: 12:46)

- Topoisomerase I is similar to topoisomerase II in that it relieves the torsional stress of supercoiled DNA during replication, transcription, and the repair of DNA.
- The difference is that it cleaves only one strand of DNA, whereas topoisomerase II cleaves both strands.
- Topoisomerase IV is a bacterial enzyme that carries out the same process as the mammalian enzyme topoisomerase II and is an important target for the fluoroquinolone antibacterial agents

Topoisomerase I is very similar to topoisomerase II. In that it relieves the torsional stress of supercoiled DNA during replication, transcription and repair. So this stress that the supercoil formed is obviously not useful when we are doing directly, if you do replication directly from the supercoil form, it is not going to be useful. So therefore the stress has to be relieved and only then replication and transcription can occur.

Now the difference between topoisomerase I and topoisomerase II is that it cleaves only one strand of DNA whereas topoisomerase II cleaves both strands of DNA. There is another isoform which is called as topoisomerase IV which is a bacterial enzyme and it carries out the same process in bacteria and is a very important target for fluoroquinolone antibacterial agents. We shall look at fluoroquinolone antibacterial agents later in the course.

(Refer Slide Time: 13:41)

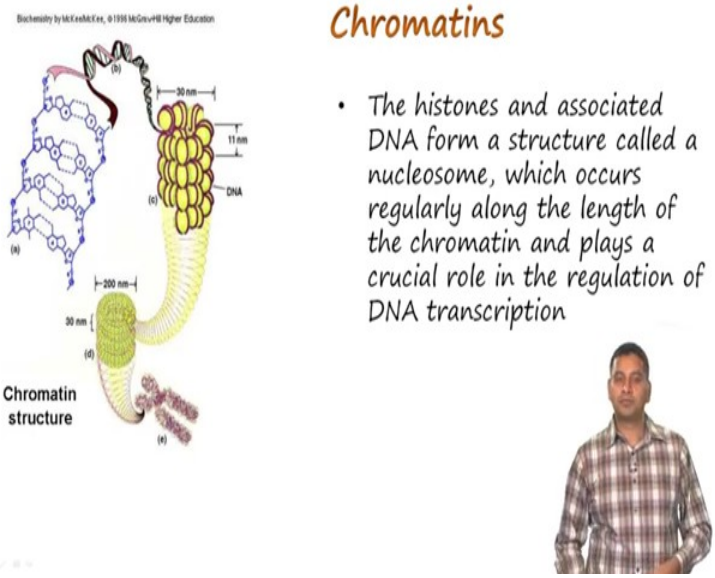


Histones

- DNA is not an isolated macromolecule within the nucleus of the cell.
- It is associated with a variety of proteins, such as histones, in a structure called a chromatin

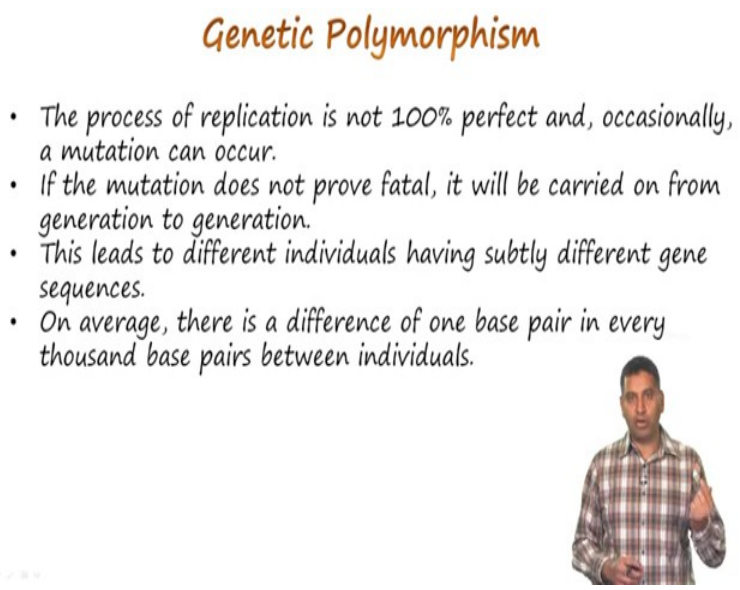
So DNA is not an isolated macromolecule within the nucleus of the cell. It is associated with a variety of proteins, such as histones in a structure called as chromatin. So as we just saw this is the primary structure of DNA where you have a sequence which can then fold into a secondary structure which is the double helix as shown here. And then this double helix can be coiled and supercoiled. And once it is, it can then be associated with proteins such as histones and it is then further packaged into histones and chromatin.

(Refer Slide Time: 14:18)



Histones and associated DNA form a structure called as a nucleosome which occurs regularly along the length of the chromatin and plays a crucial role in the regulation of DNA transcription. We shall look into some of these concepts later in the course.

(Refer Slide Time: 14:35)

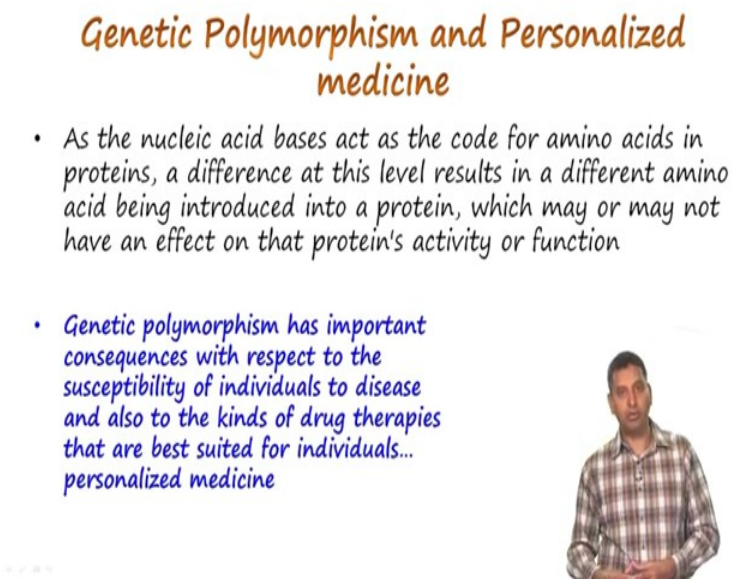


Lastly we are going to look at genetic polymorphism. So once this replication occurs, that is DNA double helix, it forms its DNA double helixes, occasionally this replication does not occur to 100 percent efficiency. That means that there are, there can be a mutation that occurs. If this mutation is fatal, that means the cell does not survive because of it, then obviously it is not

carried over to the next generation. But if the mutation is not fatal, then it can be carried from generation to generation.

What happens is that this leads to different individuals having slightly different gene sequences. Now this becomes important because when we are administering drugs to individuals, each individual can actually have a slightly different reaction. On an average there is a difference of one base pair every thousand base pairs between individuals.

(Refer Slide Time: 15:38)



Now as I mentioned earlier, this difference in sequence translates to a difference in the amino acid code or the primary structure of amino acids. So any single change in the sequence of DNA can result in a different amino acid being incorporated into the protein. This single amino acid change may or may not have an effect on the proteins structure or function. But it is important that this difference can manifest when it is interacting with certain drugs.

Genetic polymorphism has very profound consequences in the susceptibility of individuals to disease and also to therapy. So now there is an effort that is being made to incorporate these types of concepts when we are treating individuals. This is known as personalized medicine. So if the sequence of the individual is known and if you are able to map out some of these differences that occur and if you are able to predict or understand how these sequences or these differences in the protein structure can manifest itself during treatment, then it is possible to be able to avoid fatalities if a person reacts negatively or to be able to administer a drug which can

be more efficacious to the individual. So this entire field of personalized medicine is now emerging.