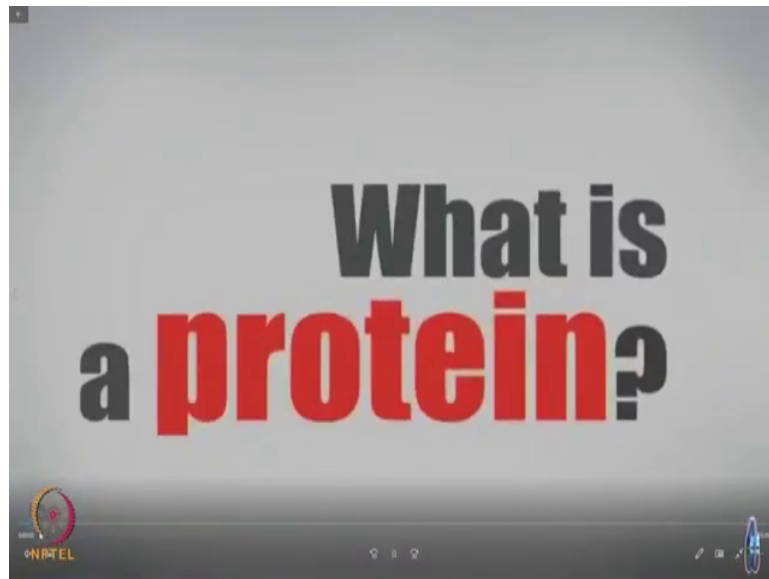


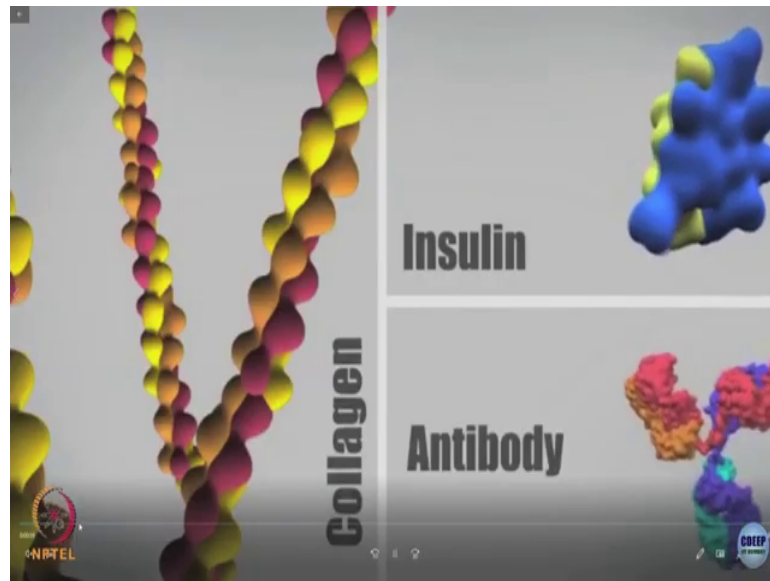
Physics of Biological Systems
Prof. Mithun Mitra
Department of Physics
Indian Institute of Technology, Bombay

Lecture – 03
The relationship between shape and function of biomolecules

(Refer Slide Time: 00:17)

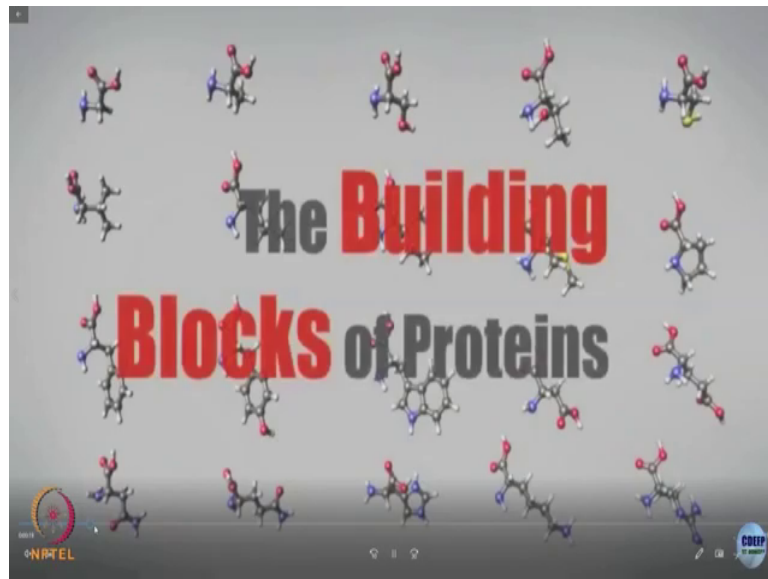


(Refer Slide Time: 00:21)

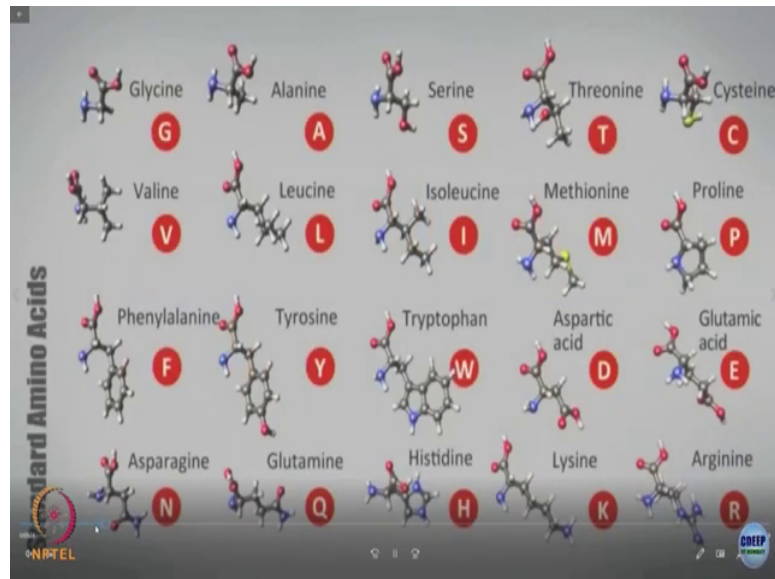


So, you can divide these amino acids into different classes.

(Refer Slide Time: 00:29)

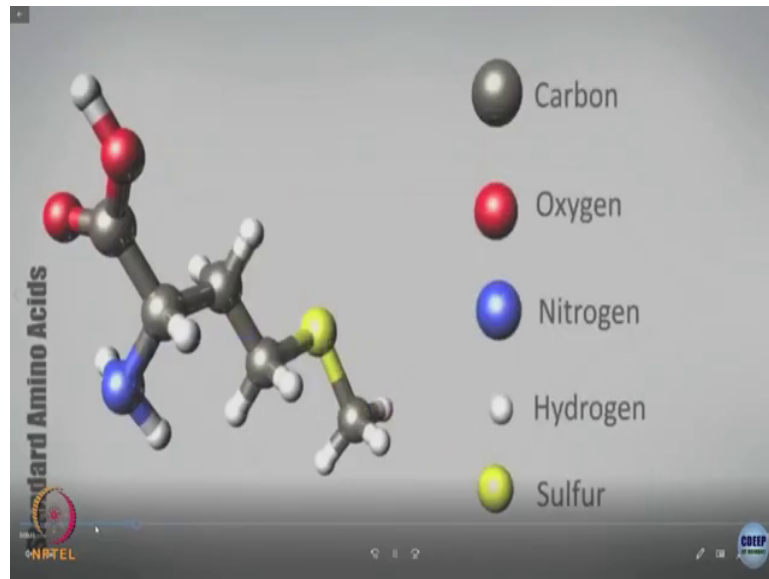


(Refer Slide Time: 00:31)



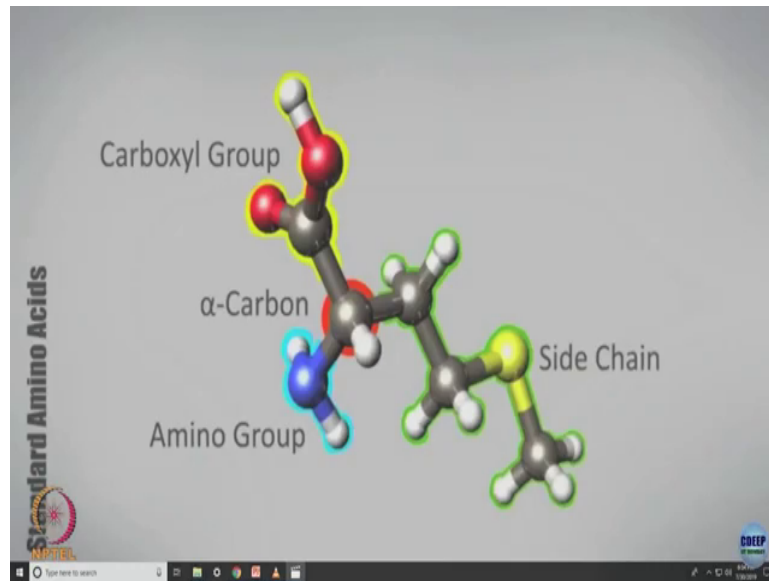
So, these are my 20 amino acids, glycine, alanine, serine, threonine and so on and so forth.

(Refer Slide Time 00:39)



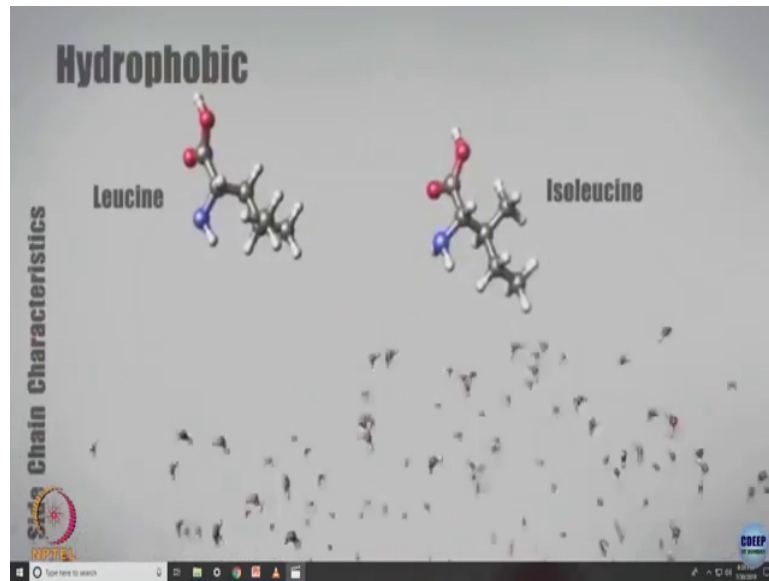
You can sort of characterize the fundamental properties. So, one class of amino so it is made up of carbon, oxygen, nitrogen, hydrogen, sulphur.

(Refer Slide Time 00:45)



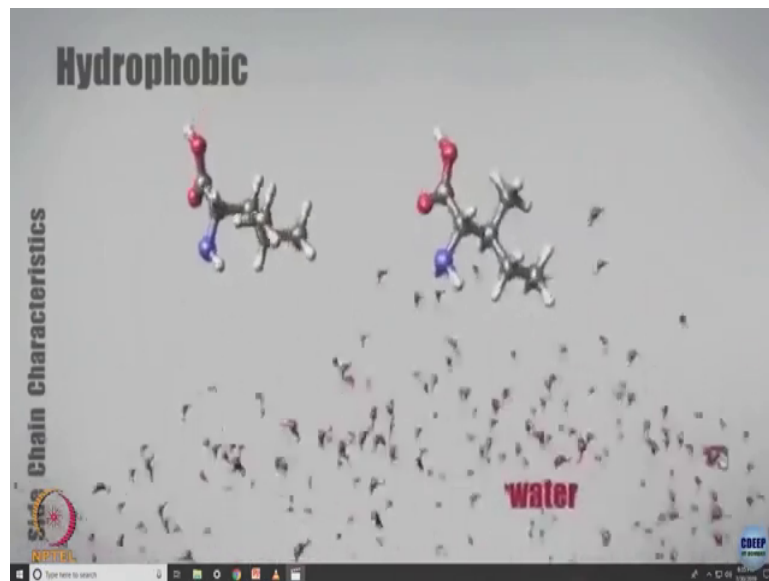
This over here is my side chain which; so, this is the fundamental the main chain and this is the side chain.

(Refer Slide Time 00:56)



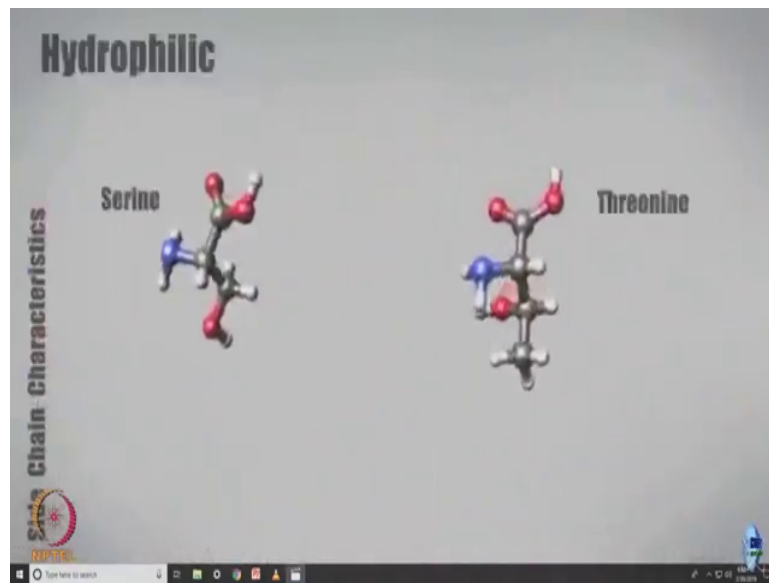
The side chain distinguishes the different amino acids.

(Refer Slide Time 00:57)



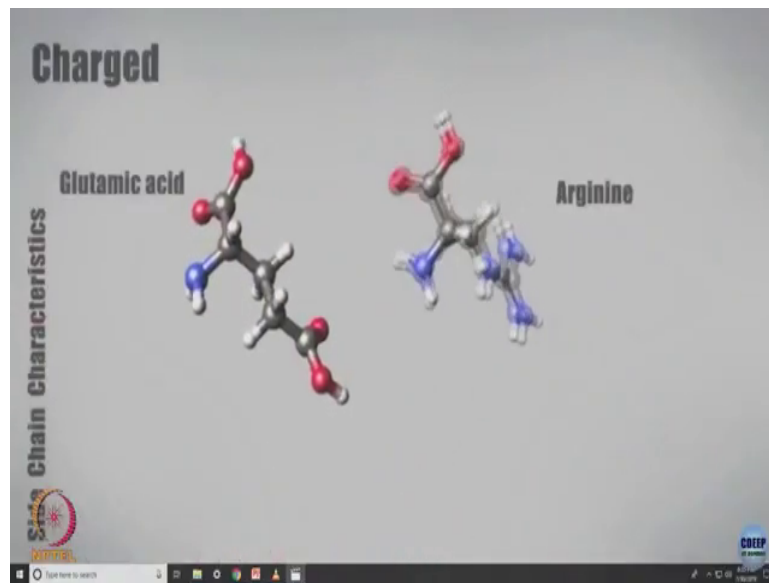
There are a few classes. So, one class is the hydrophobic amino acids, which do not like to interact with water.

(Refer Slide Time 01:04)

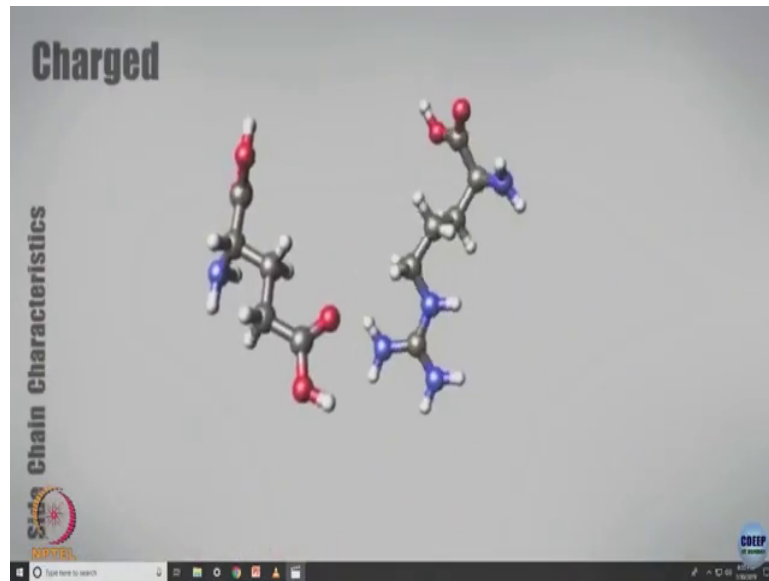


So, whenever you have hydrophobic amino acids they will try to tend to cluster together.

(Refer Slide Time 01:12)

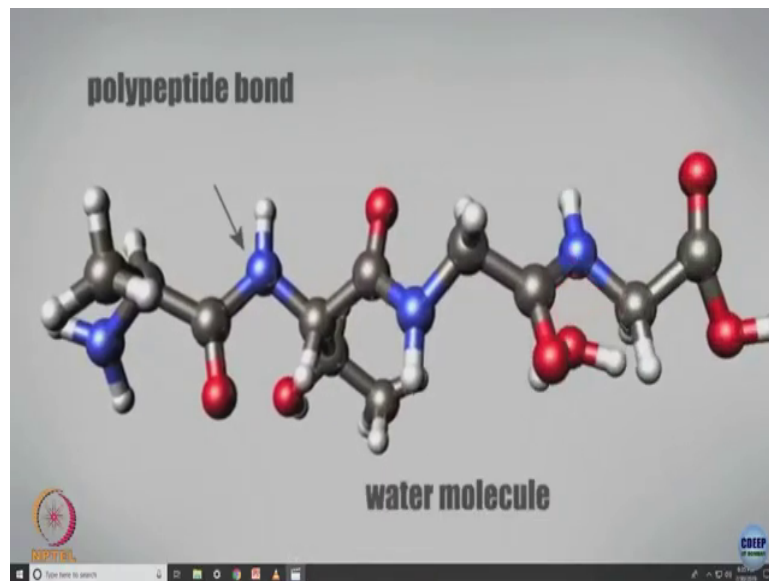


(Refer Slide Time 01:14)

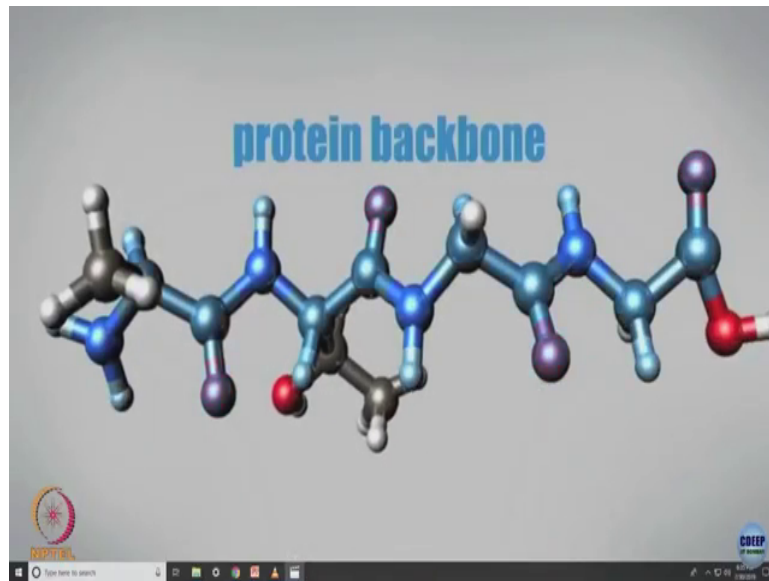


You could have hydrophilic amino acids which like water and therefore, if you put in hydrophilic amino acids they will try to sort of stay away from each other and closer to water.

(Refer Slide Time 01:21)

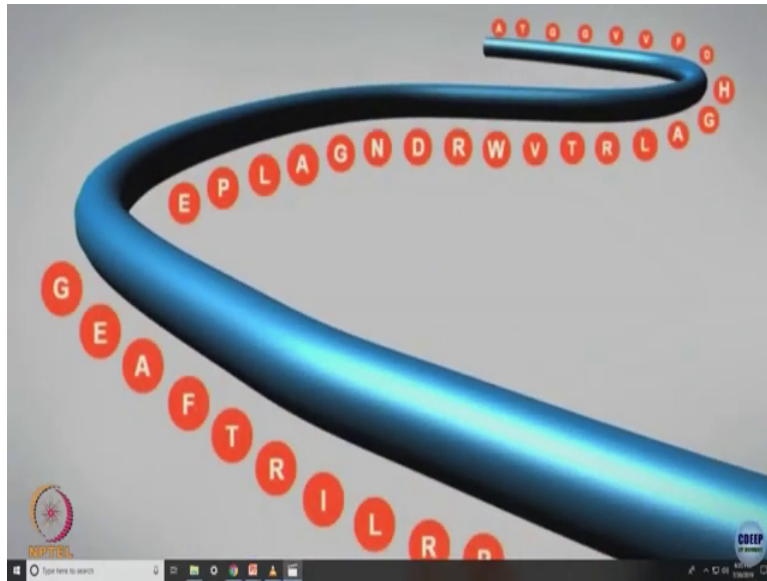


(Refer Slide Time 01:26)



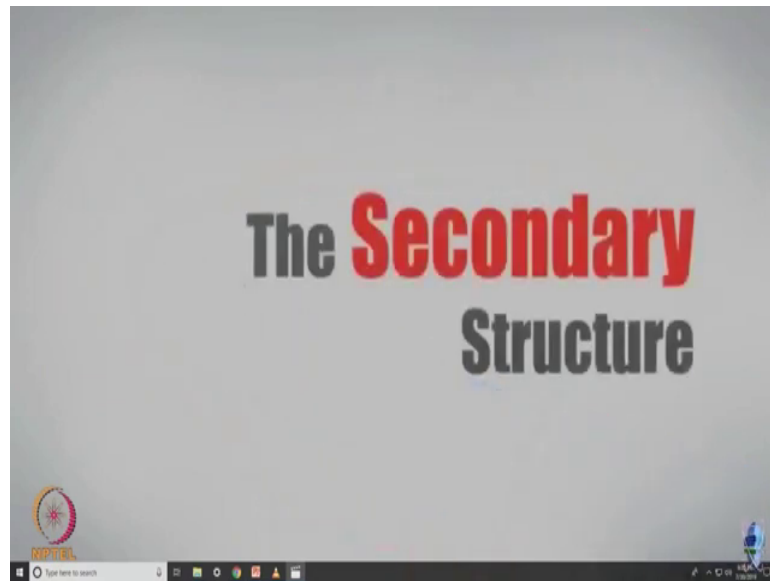
Similarly, you can have charged amino acids and I will upload them which ones are charged and which ones are not.

(Refer Slide Time 01:29)

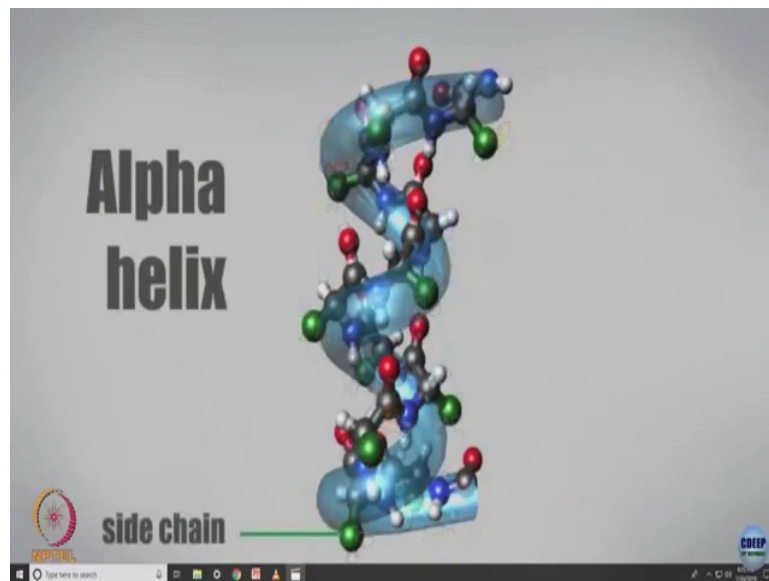


So, you can take these amino acids and join them together to form what is the protein backbone. So, here is my sequence of amino acids that forms this protein backbone.

(Refer Slide Time 01:36)

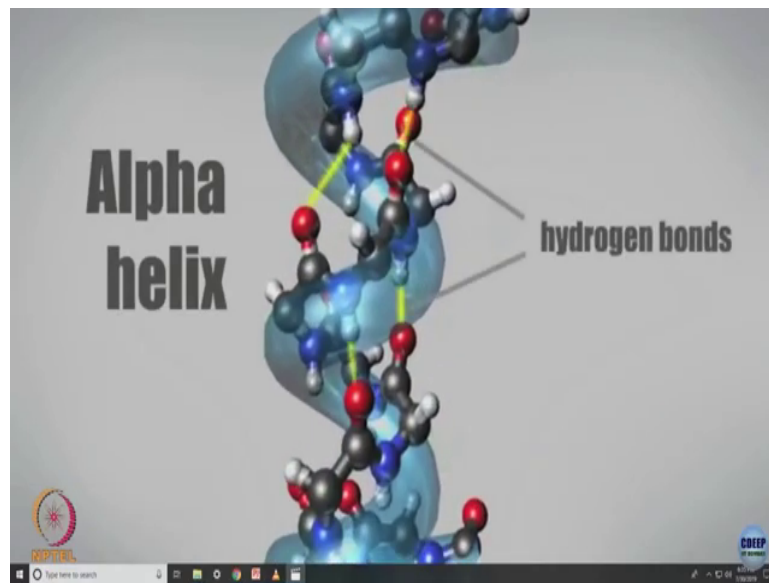


(Refer Slide Time 01:38)

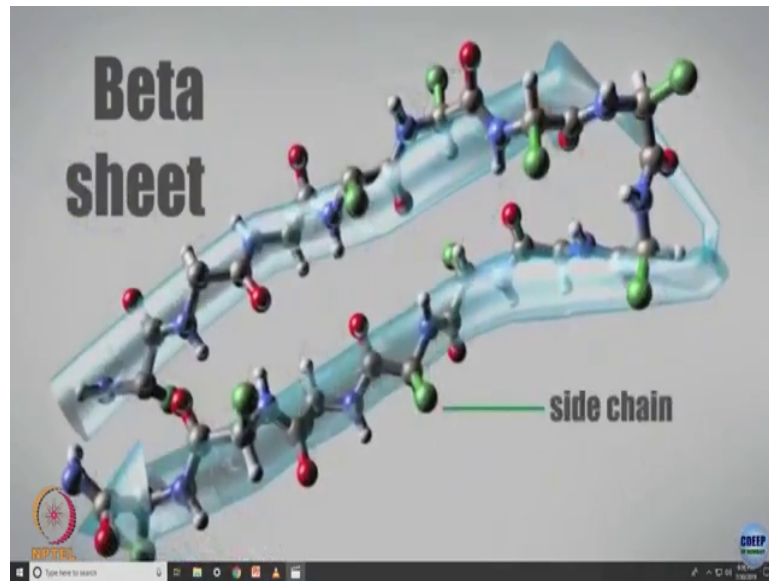


Given a sequence the protein can fold in different ways so that first level what is called the secondary structure, you can have alpha helices and beta sheets. Depending on the sequence of amino acids and whether correct hydrogen bonding can form or not.

(Refer Slide Time 01:47)

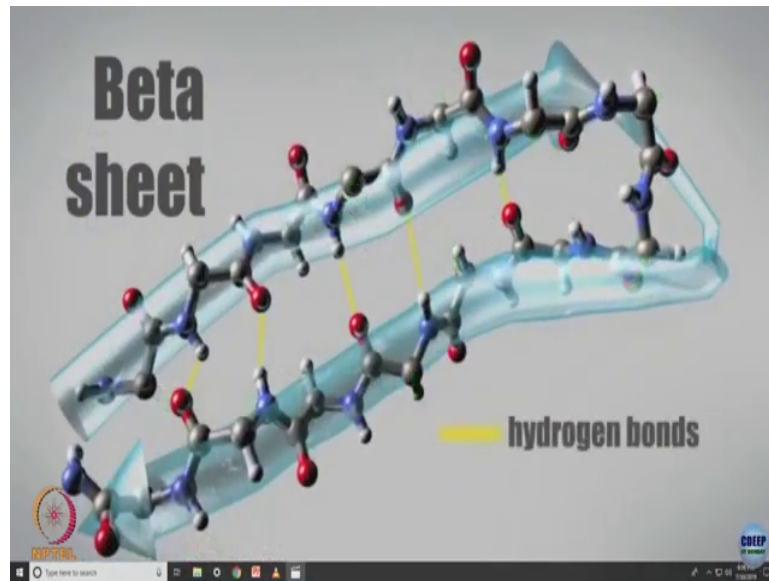


(Refer Slide Time 01:52)



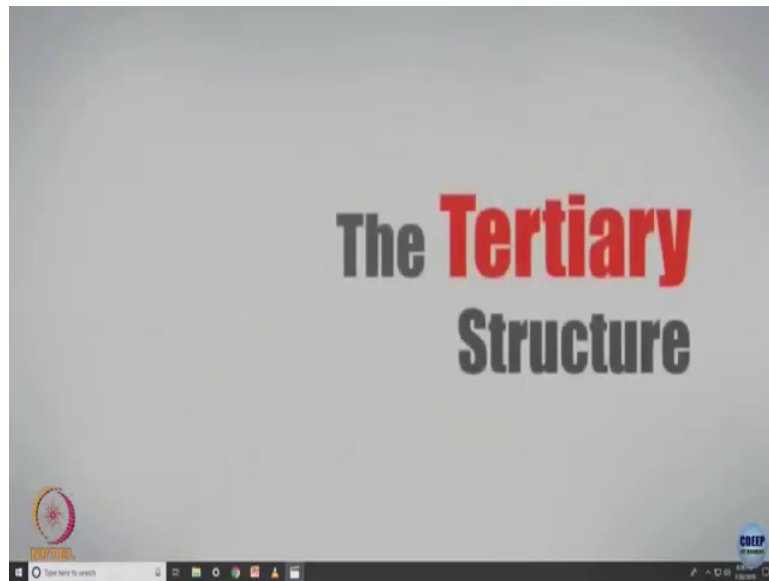
So, you have this alpha helix and the beta sheets.

(Refer Slide Time 01:56)



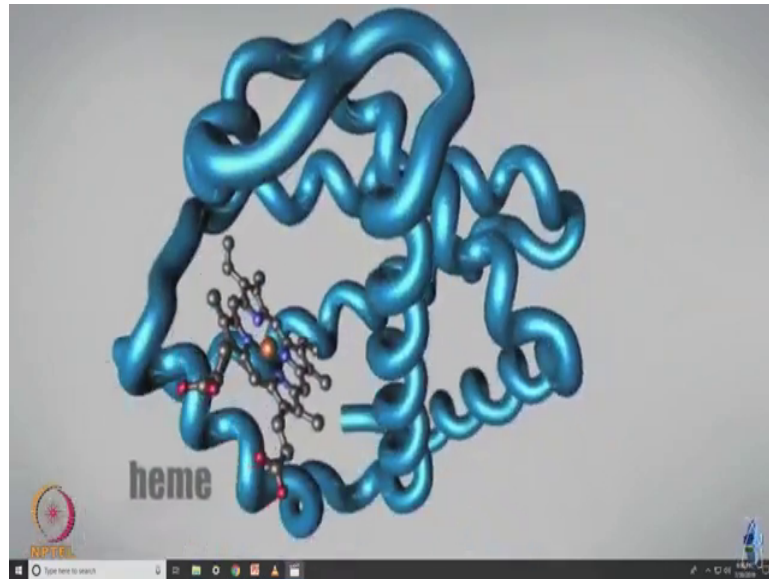
So, these are called the secondary structures.

(Refer Slide Time 01:59)



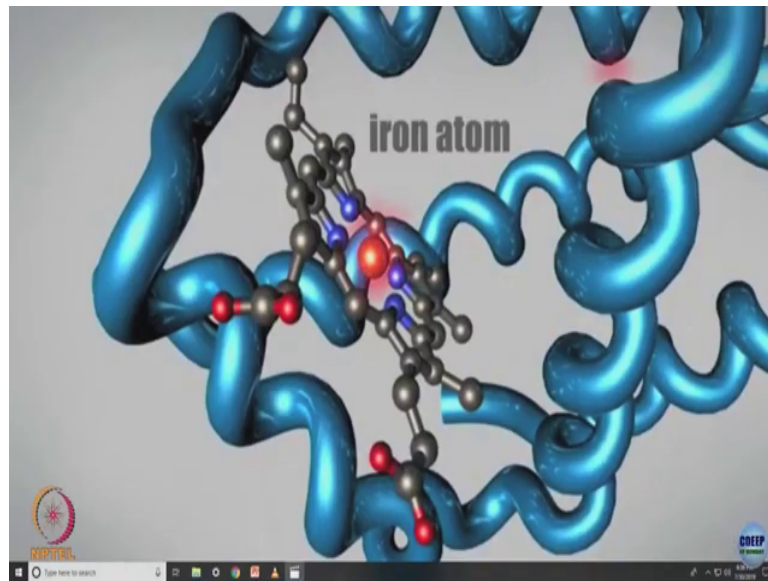
You can have a high-level structure which is called the tertiary structure.

(Refer Slide Time 02:02)



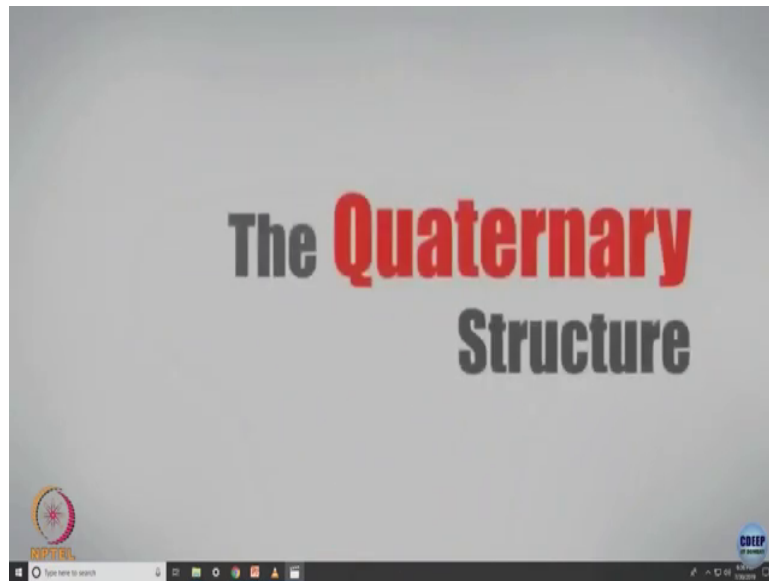
So for example, these you have alpha helices over here which coil up amongst themselves.

(Refer Slide Time 02:09)

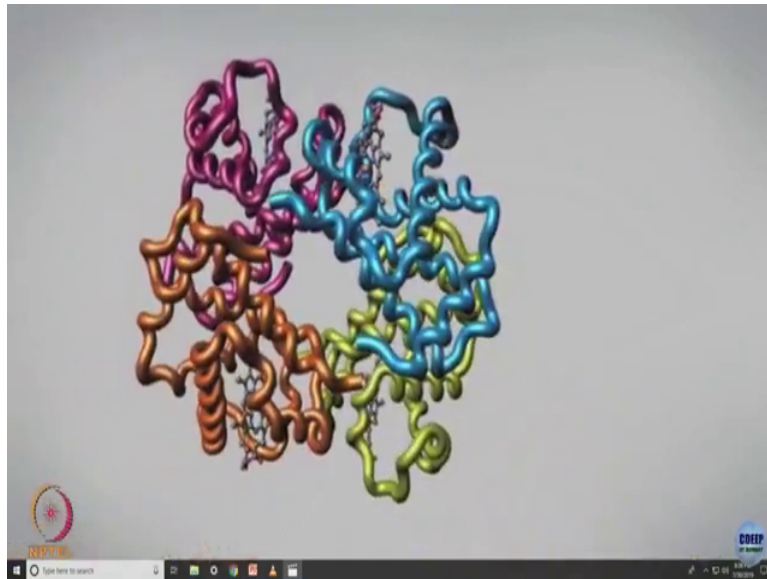


So, this is for example, the haemoglobin protein. So, depending on how it folds, so this is the protein folding.

(Refer Slide Time 02:16)

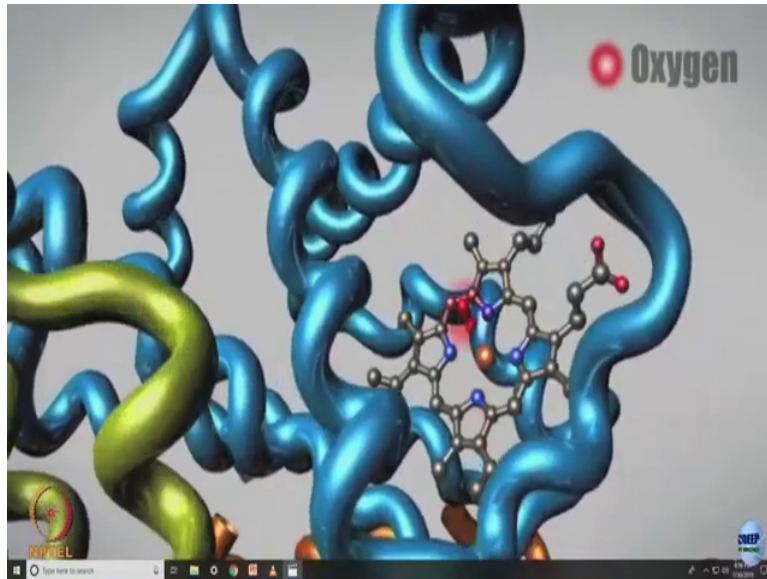


(Refer Slide Time 02:20)



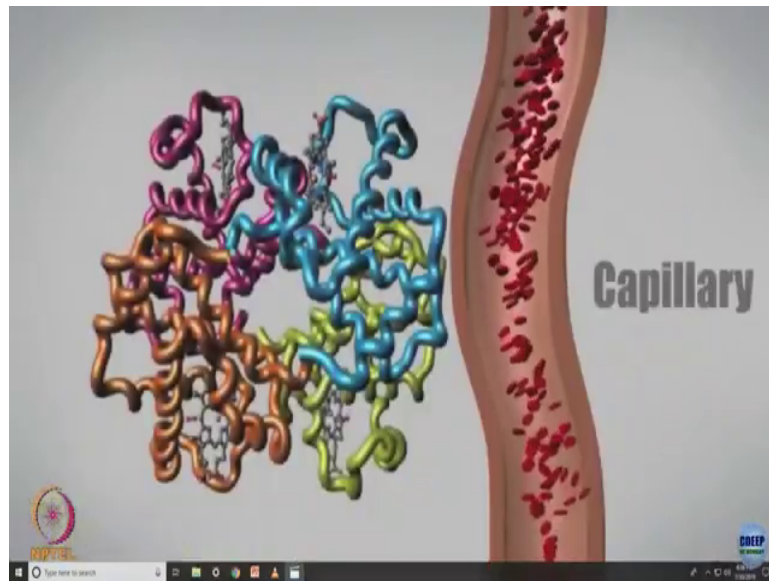
It determines its function for example, the hemoglobin forms a pore where the heme atom can come and bind.

(Refer Slide Time 02:23)



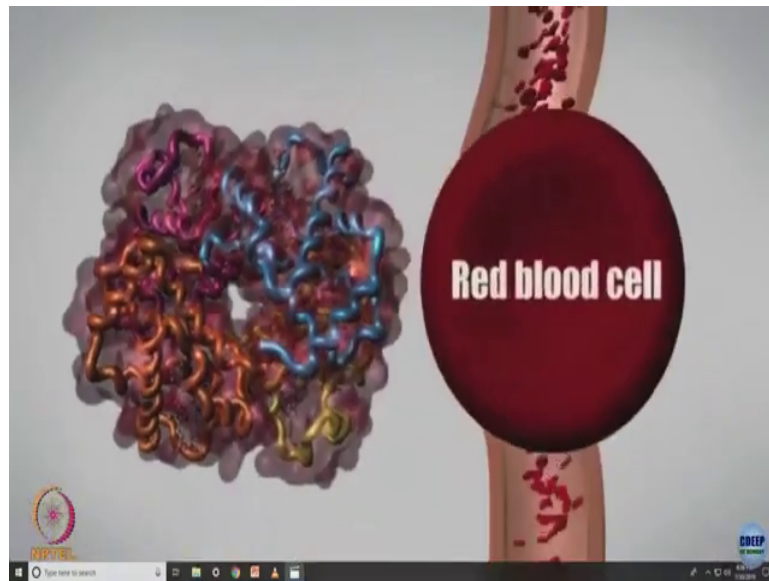
And then there is the quaternary structure where different polypeptides can come to.

(Refer Slide Time 02:26)

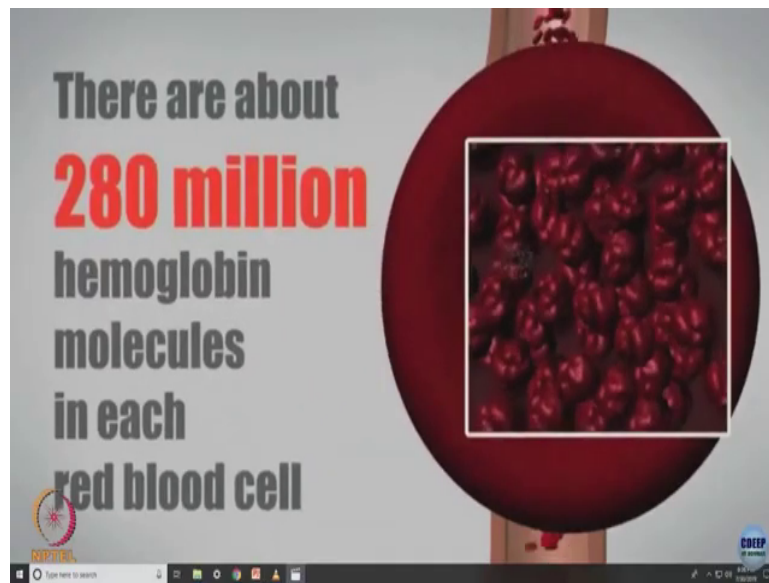


So, hemoglobin contains four of these heme oxygen binding sites.

(Refer Slide Time 02:29)

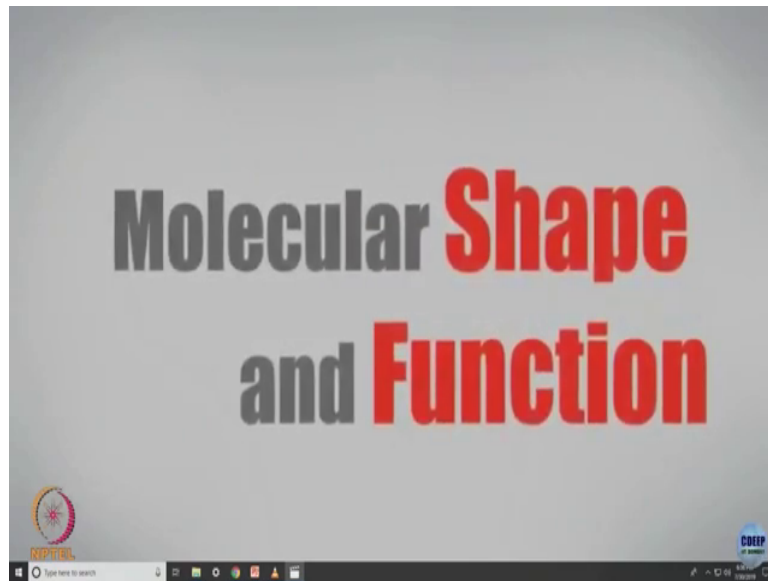


(Refer Slide Time 02:31)



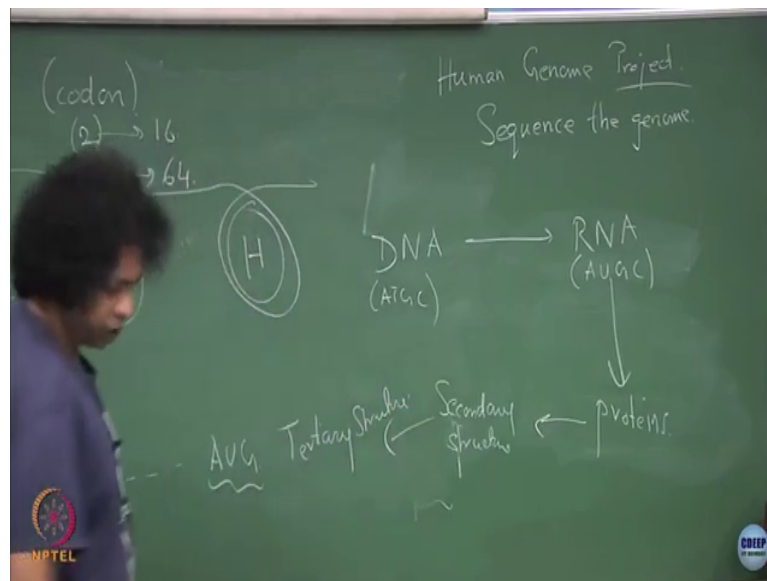
And together you have this hemoglobin protein, and there are an immense number of them.

(Refer Slide Time 02:36)



So, let me that us all I want to say as far as proteins goes. So, you have the primary sequence for proteins.

(Refer Slide Time 02:48)



So, the sequence from there on you have the secondary structure which are alpha helices and beta sheets and from there on you have the tertiary structure which is actually the folded structure of the protein and so on. So, when we go on to when will talk about protein folding later on in the course, we will sort of use these properties.

So, again protein folding is sort of an open question right. So, that is one of the most common themes that I will tell again and again in this course no question in biology is completely solved everything is an open question right. So, that is why it is nice because you have lot of areas in which you can contribute, but nothing is completely understood.

So, protein folding is of course, is all the canonically hard problems to solve, but the way one at least starts to think about them is to try to group these amino acids depending on their properties depending on so, the simplest class of modelling depends on whether with the

amino acid is hydrophobic or hydrophilic or charged. And depending on that you try to build model which are called these hp models of Dylan Thomas and company anyway. So, I will talk about that when I get to protein folding, at least a little bit about that not allowed I have another videos.

(Refer Slide Time 04:15)



So, here is the sort of central dogma what I was talking about the transcription translation process, it is actually a fascinating thing. So, even if you know or if you have forgotten it is nice to know once again.

(Refer Slide Time 04:18)



So, here is my DNA strand right, it contains this ATGs and C s, I have this machinery of proteins that comes and binds to this DNA strand. It unzips the DNA strand which means it opens up this double helix and makes a single strand ok. So, once you have made a single strand this machinery sort of zips along this DNA and it reads off this basis and it produces the MRNA.

(Refer Slide Time 04:45)



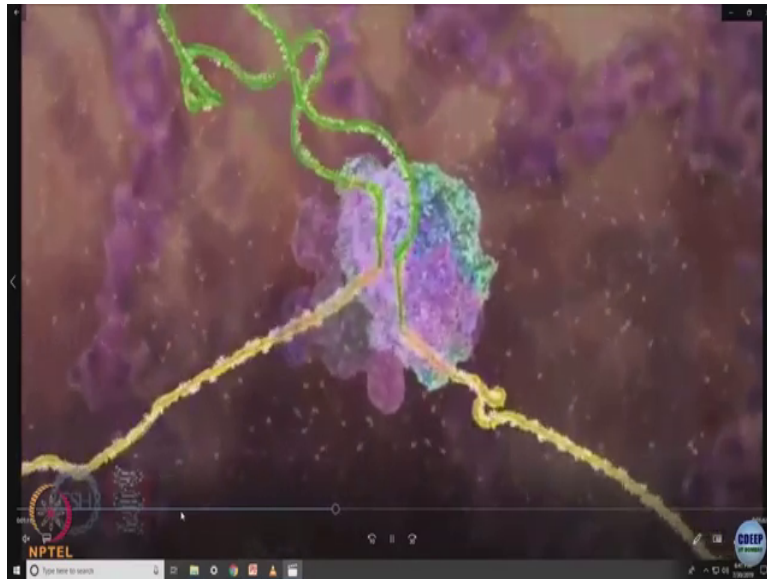
So, this yellow thing is my mRNA which is being produced. So, let me pause ones, but whatever. So, here in this DNA its where you have this ATGs and C s on this yellow ribbon which is my RNA you have this a UCGs and Cs ok. So, that is where the uracil comes in. So, it produces a copy of this DNA in this RNA molecule ok.

(Refer Slide Time 05:13)



Now, the DNA contains introns and exons which means that only some parts of this are required to actually produce a protein.

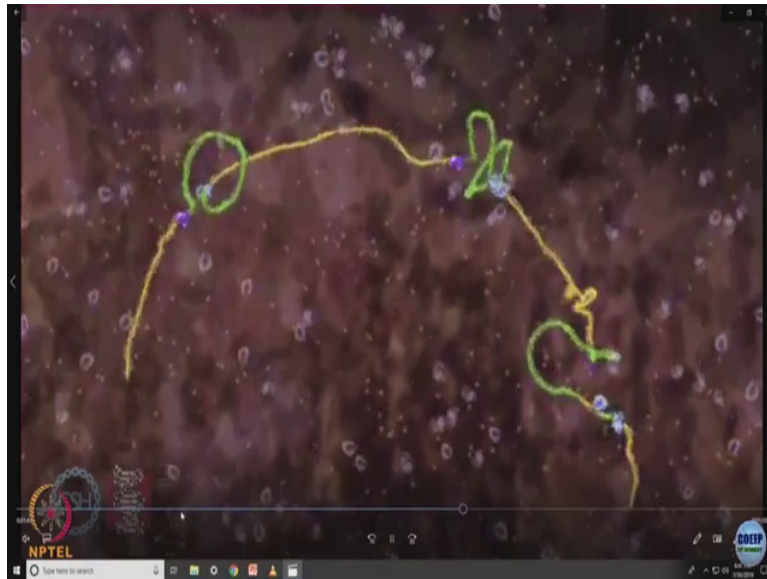
(Refer Slide Time 06:28)



So, what the cell will do is that will splice off those parts using another protein machinery which is called a spliceosome. So, here this thing the spliceosome will form, it will bring together the two ends of this green strand and will cut them off it will join the yellow parts to one another.

So, that the x trans are connected together ok. So, it is a fascinating thing how the cell does this sort of reproducibly again and again and again and again we do not understand the complete mechanics of this. So, it will splice off that green part at some point and join the exons yes that is it. So, it will do that over the length; over the length of the gene.

(Refer Slide Time 07:04)



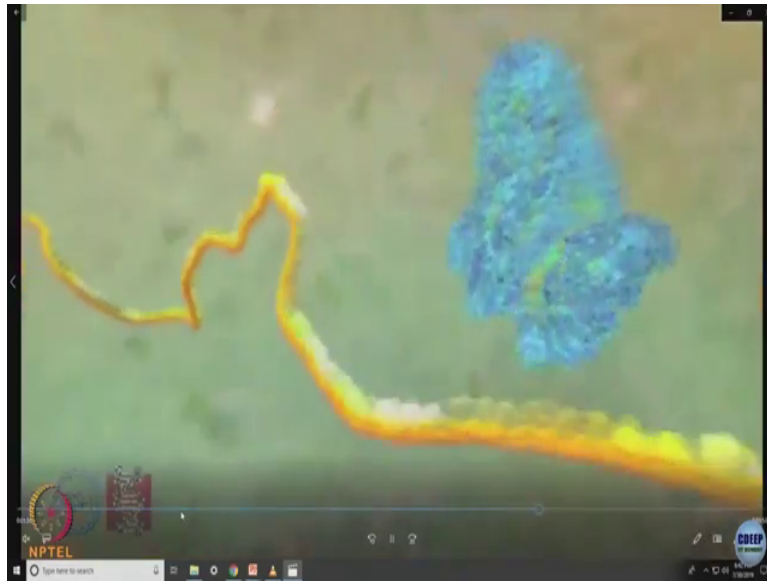
So, gene typically will have many introns and exons. So, it will strip off all of these green regions and get you a continuous yellow strand which is all the coding part of the DNA coding part of the gene the green parts the non-coding parts ok. So, all those green parts of now been cut off and you have this polymer which is like this whole coding yellow polymer.

(Refer Slide Time 07:28)



So, this yellow polymer is then ejected out from the nucleus.

(Refer Slide Time 07:32)



So, it comes out of the nucleus out in the cell cytoplasm.

(Refer Slide Time 07:36)

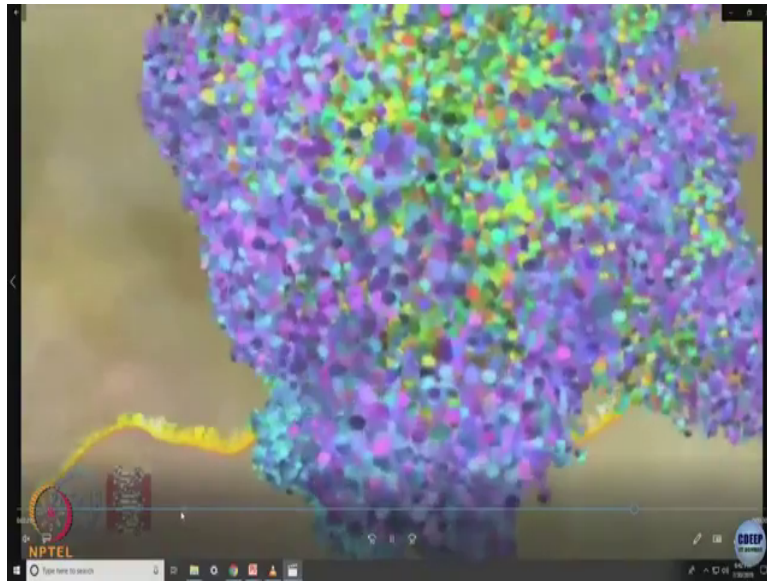


You have this another protein machinery which is called the ribosome, the ribosome then reads this mRNA this messenger RNA this yellow thing that came out and it constructs the protein.

(Refer Slide Time 07:51)

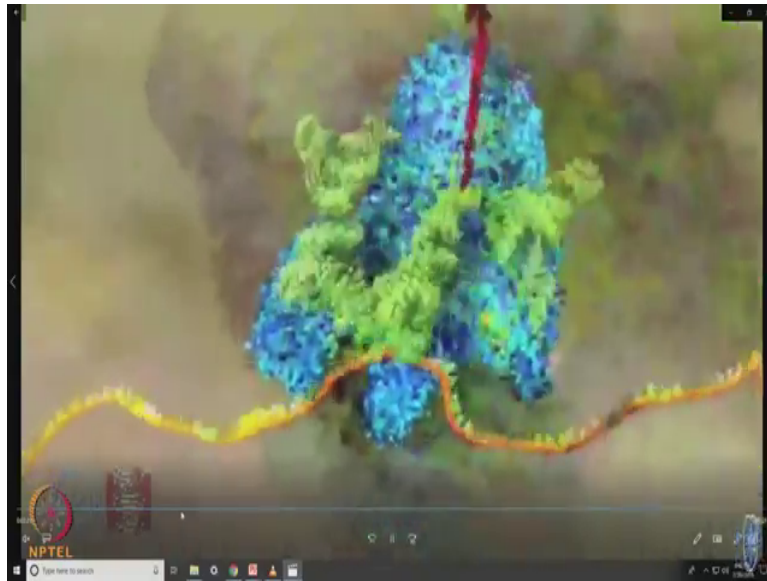


(Refer Slide Time 07:54)

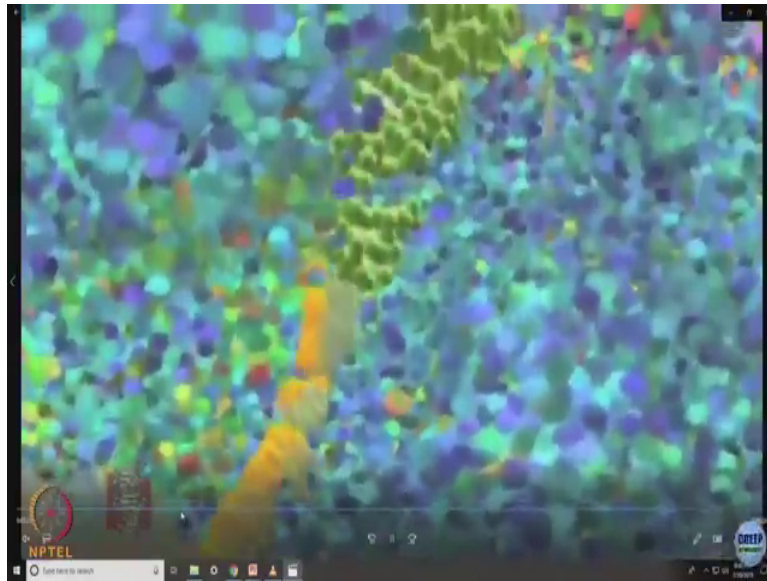


So, these green things are transfer RNA's which are bringing these amino acids, corresponding to these sequences and as it brings in these amino acids it constructs the protein one by one by one or codon by codon by codon.

(Refer Slide Time 07:57)



(Refer Slide Time 08:03)



And this red thing comes out which is your protein. This red thing will ultimately fold and form the tertiary structure of the protein ok.

(Refer Slide Time 08:08)



So, this is the central dogma transcription DNA gives rise to RNA gives rise to proteins; so, this transcription translation process. So, here is the protein molecule coming out of the ribosome and will fall fold into whatever structure it is suppose to fold in for its particular function. And the same ribosome will sort of read any different RNA's and produce the right protein, that is encoded in the sequence in the MRNA.

Student: (Refer Time: 08:36).

Here, or when in the first part or in the second part?

Student: (Refer Time: 08:45).

Not here.

Student: (Refer Time: 08:49).

This one.

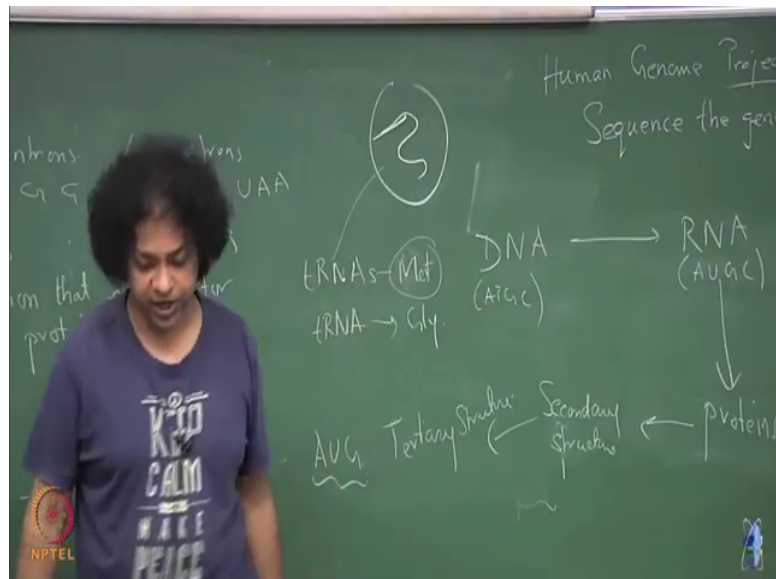
Student: Plate.

Plate.

Student: (Refer Time: 08:54) background .

Those green things, maybe the is really those are the transfer RNA's this thing right.

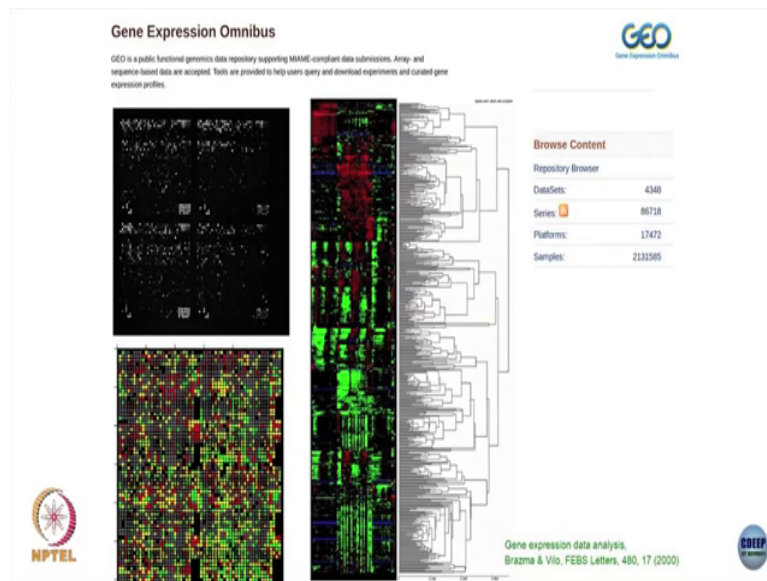
(Refer Slide Time 09:07)



So, these are the transfer RNA's, these tRNA's contain an amino acid. So, they will have let us say a methionine or another tRNA will have let us say glycine and so on. So, corresponding to each amino acid you have this package which this transfer RNA which will bring this correct amino acid ok.

So, once it brings it into the ribosome, the ribosome takes off this methionine and it releases the transfer RNA back. So, these are sort of your Amazon deliveries which are bringing the right package so, that you can incorporate them according to the sequence into the protein right.

(Refer Slide Time 09:45)



Now, as I said so this part at least we understand how this happens, the DNA produces RNA produces proteins, but like I said the cell contains many many many genes right, genes are regions of the DNA. So, what is a gene? A gene is a region of the DNA which encodes for a protein, region that encodes for a protein that encodes for a protein ok. So, of course, you have many many millions of genes, it is not necessary that all of these genes are going to be active at the same time right or in lead active in the same cell type.

So, how these genes are expressed itself is a field of study. So, these are experiments done on DNA microarrays, where you so each spot in this microarray is a single gene and then the intensity of this fluorescent spot tells you how active that gene is. You can do that, so you can do this. So, this is this is a black and white microarray which you can convert to a colour and

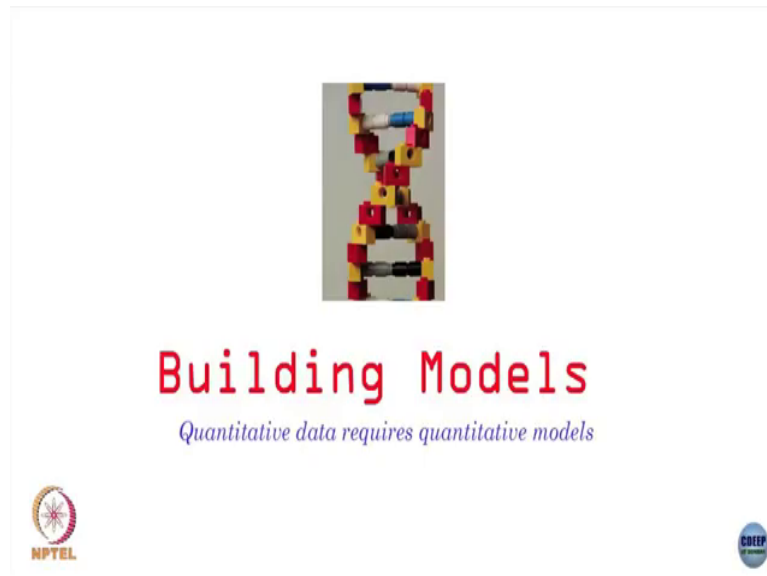
then you can do that over a over a period of time. So, this axis is time and you can do that for different genes.

So, in this picture over here my pointer does not work, but in this long picture over here the x axis is time the y axis is different genes ok. So, different and I think if I remember correctly green is sort of strongly on and red is strongly off, I could be mistaken it could be the other way round. So, you see that it gene stays on for some time then it switches off and then again it turns on back at some other time. Something stays off it switches on then again turns off and you can look at how these how these gene expressions are correlated in time. So, which genes switch off with what other genes ok.

So, you can build these correlation matrices among genes by looking at this sort of gene expression data as a function of time and that again tells you functional information about how genes crosstalk with each other, how different proteins interact with each other. So, if one protein is interacting with another protein those two genes might be turned on at the same time or they may be turned off at the same time ok. So, these are enormously complicated, but again this data is available, we do not exactly completely understand this data of course, but at least as far as modelling purposes goes a lot of data is available.

So, again this gene expression omnibus is again an open source genomics database where you can go in and you can see what sort of this gene expression data's for different organisms for different cell types are available. And again these are slightly older numbers I am sure the numbers have sort of improved in the time sense I drove this slide alright.

(Refer Slide Time 12:43)



So, the idea so now that I have talked a little bit about biology. So, what I hope to achieve during this if you do by this course is that we cannot cover all of this biology that is a given biology is enormously complicated I will cover maybe 5 percent right. But, what I hope that you will take away from this course is an idea of how should you given a biological problem that you are interested in, how should you go about building a model that takes into account the quantitative data that is available and makes quantitative predictions ok.

(Refer Slide Time 13:20)

Overlook complexity, generate abstraction, develop mathematical model

↓

Develop intuition and insight, generate quantitative, testable predictions

HOW TO BUILD A MODEL?

- ▶ Must satisfy the basic physical laws (Conservation laws, laws of thermodynamics, ...)
- ▶ What are the relevant features? What are the relevant spatial and temporal scales?
- ▶ Correct physical descriptions at the relevant scale
- ▶ Write down mathematical equations to simulate the system
- ▶ Generate testable predictions
- ▶ A wrong model can sometimes teach us as much as a right model!

NPTEL

CEED

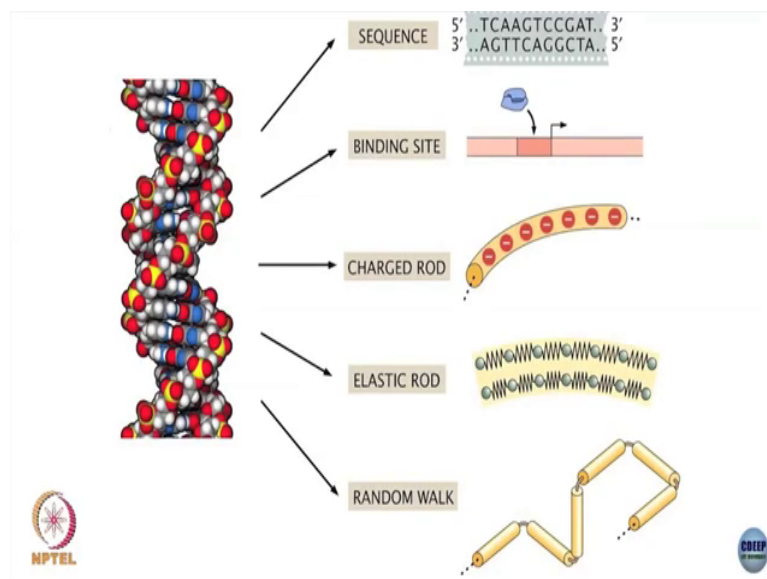
We will do different systems, but hopefully the idea the ideas that will develop will be applicable across whatever other system that you might be interested in. So, the generic idea is that of course, you sort of try to identify which parts of the system are relevant to the question that you are asking. And then you overlook the other parts the other complexities and you develop a mathematical model and through that model you generate quantitative testable predictions right.

What should this model satisfy? Of course, it should satisfy basic physical laws, just because biology is complicated does not mean it can violate physics, you will have whatever conservation laws you will have whatever laws of thermodynamics that we know. The key thing is to sort of identify what are the relevant features for the question that you are trying to ask.

So, if you are trying to ask about this chromosome folding, does it really matter what exactly the sequence is. Maybe it matters maybe it does not I do not know, you might say that well because this is at a scale which is one-third of a nanometer and this packaging is at a scale which is of a micrometer. Maybe the exact sequence details do not matter. You may be right, you may be wrong I do not know, but that is one of the key things that one needs to learn how to identify, that what the relevant features, what are the relevant spatial and temporal scales that I should be talking about.

Once you have identified that you can guess what is the correct physical description and write down these mathematical equations, once you have a mathematical equation you can generate predictions and then if you are right great. If you are wrong you tell yourself that you learn something which is true and you go back and generate a new better model hopefully ok.

(Refer Slide Time 15:08)



Just to show you what I mean by this identifying the correct scales and the correct features of the problem let me give one or two examples. So, here is DNA right so which we have been talking about. So, at the very microscopic level if I go here closely into the DNA what I have is this sequence right over there in the top and you might. So, if you are asking questions at a very microscopic level that will a protein come and bind here or there that depends exactly on the sequence of the DNA.

So, you need to know the DNA you need to know the DNA sequence if you need to make any predictions of that sort. You can maybe abstract a little bit and say that well I just abstract this sequence into two types of sites a binding site or a non-binding site for the DNA and then this protein might have different maybe it has thousands of binding sites over the genome.

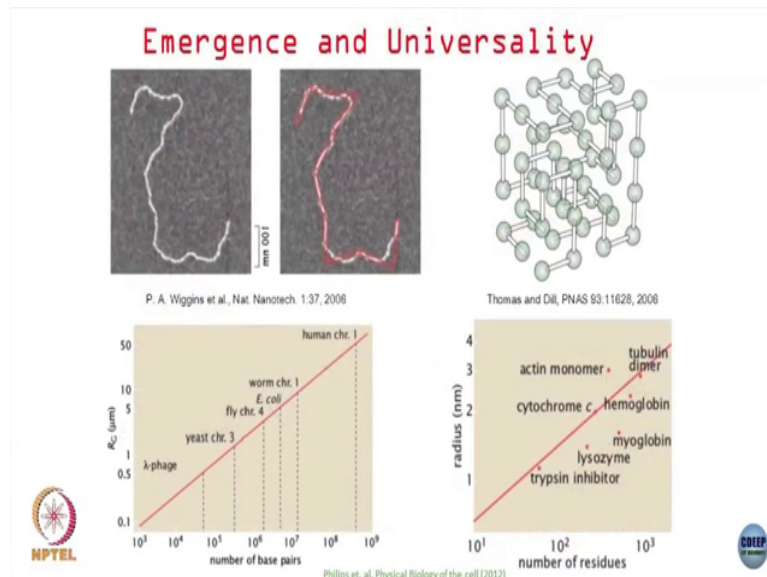
So, instead of saying that its ATG's and C's I will just convert this into a binding site a non-binding site a binding site and so on ok. That is another level of that is level of abstraction, but if you are just interested in how much a protein will bind maybe that is enough. If you zoom out a little more if you are not interested in such granular microscopic details you could maybe treat the DNA as a charged rod.

So, these nucleotides have negative charges exposed. So, you can treat the DNA as a negatively charged helix double helix or because it is a very long object you could treat it as a sort of elastic rod. So, beads connected by springs and then you could ask how easy is it to bend the DNA. So, for example, because you have to package the DNA inside the nucleus of course, the DNA cannot be very stiff, you need to be able to bend the DNA like this, but at what scales can you bend it.

So, what is the persistence length of the DNA? Questions like these maybe you do not need to know the exact sequence maybe you can just model it as a bead. And spring and then depending on the stiffness of the spring, if you get the right stiffness you might be able to capture the correct physical properties of the DNA ok. If you zoom out even more you can maybe talk of the DNA as a random walk in space and that would get some things right that would get some things wrong.

So, even if you think of DNA as a random walk it is not that that is a completely. So, good random walk in space, it is not that that is a completely bad model let me just show you this.

(Refer Slide Time 17:32)



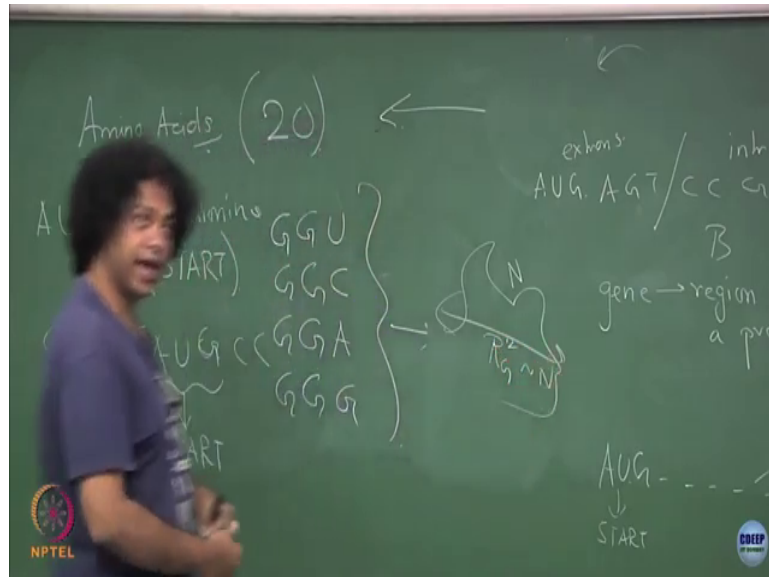
So, some of you have done random walks right. So, how does the end to end distance scale for a random walk? With the number of base pairs or with time anyone? So, if I have a random walk made up of n. So, n legs so 1 2 3 4 like that then how would the end to end distance scale with the number of units?

Student: R square (Refer Time: 18:04).

R square would go as n to the power of 1 right. So, if you look at various DNAs for various organisms. So, this is the radius of gyration which is closely related to the end to end distance, you will see that it grows as n for various different organisms. So, this is a why this

lambda phase this is east this is a fly chromosome and then there is the human chromosome over there.

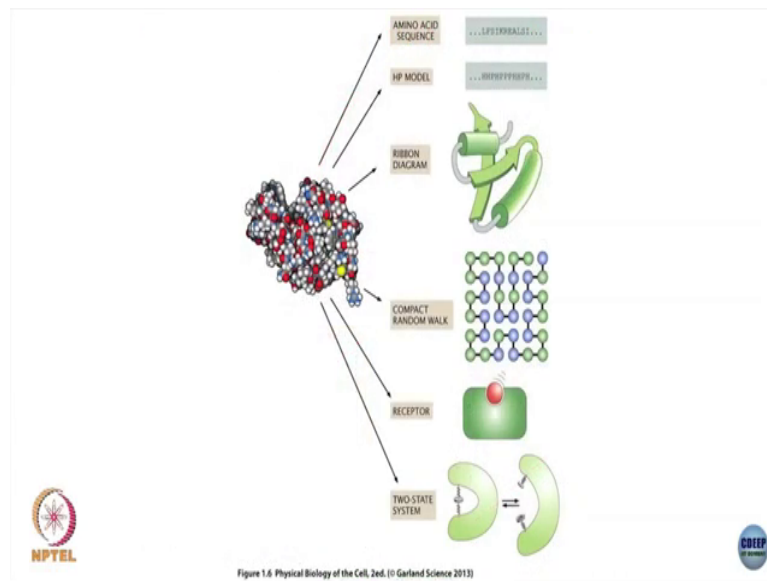
(Refer Slide Time 18:36)



If you connect all of these dots so if you just say that if I take this DNA polymer and I ask how large how large is this polymer can I get an estimate of the size of this polymer given the length of this DNA which is n it turns out that rg square goes roughly as n and that is true across organisms.

So, depending on what the question that you are interested in different levels of modelling may be appropriate, you need not have the exact microscopic details in order to ask a very macroscopic coarse grained question ok. So, you need to identify what spatial scales or what temporal scales are important in order to answer the question that you are asking and then do your modelling at that correct scale.

(Refer Slide Time 19:17)

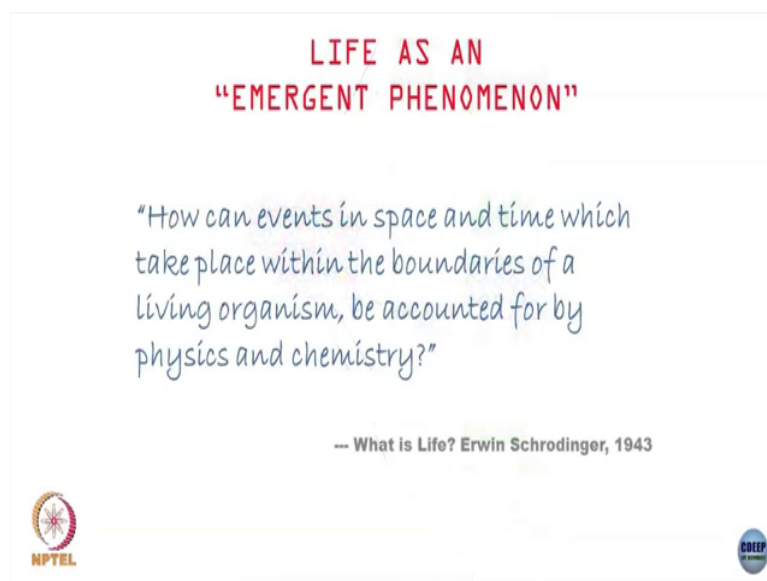


So, this is true I showed for DNA, it is also true for proteins. So, again at the most microscopic level you have your amino acid sequence which is whatever amino acids that you have. You could then sort of abstract a little of the complexity and say that I will just say whether it is a hydrophobic molecule or a hydrophilic molecule and then you come to these class of models that are call this hp models.

You could abstract every little more and say I will only talk about the secondary structure whether there is an alpha helix or a beta loop or you could just say that well I will just talk about the functionality that its maybe binds to some other protein. So, its acts as a receptor or the ligand and then I can just talk about the energetics of this binding this ligand receptor binding and again these are all of these are things that will sort of look at in greater detail as the course goes on.

But the idea again is this that the most crucial thing when doing modelling of a biological problem is to sort of identify which features are important to that problem and which features are not. If you can correctly identify that you can then go about building the correct model for your problem.

(Refer Slide Time 20:24)



So, we sort of think of biophysics as a slightly modern subject. In some sense it turns out that you know people have been thinking about it for a long time. For example, Schrodinger of Quantum Mechanics in his What is Life? Book said that thought about these same problems and I sort of crystallized the question by saying that see these are events in space and time which are taking place within the boundaries of a living organism.

And how can I account for these processes that taking place in space and time which cannot violate my laws of physics as I know it, how can I account for these processes using the tools

of physics and chemistry ok. So, that sort of the spirit in which will approach this course. That how do I take this complex processes that are taking place inside living organisms and how do I account for them using whatever physics knowledge that I have or whatever chemistry or maths knowledge that I have ok. So, that is yes, that is all for today. So, we will reconvene back here on Friday.