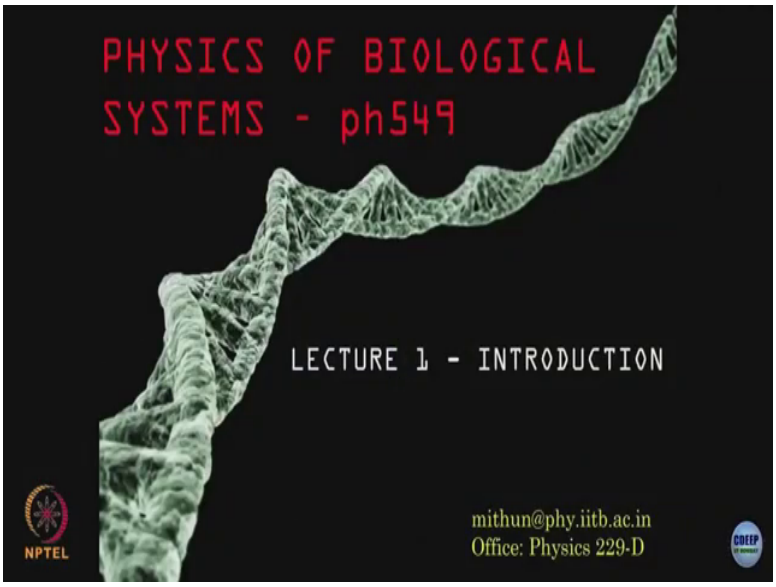
Physics of Biological Systems
Prof. Mithun Mitra
Department of Physics
Indian Institute of Technology, Bombay

Lecture – 01
Introduction to Physical Biology

We will try to get, how many of you have done anything beyond class 12 biology? Anyone? You have done something? BV 101, yes you remember anything of BV 101? Ok good. So, we will try to recap even if you have not, we will try to sort of get familiar with the lingo as it were of biology. Some common terms we will be using again and again some of which are familiar some of which may not be.

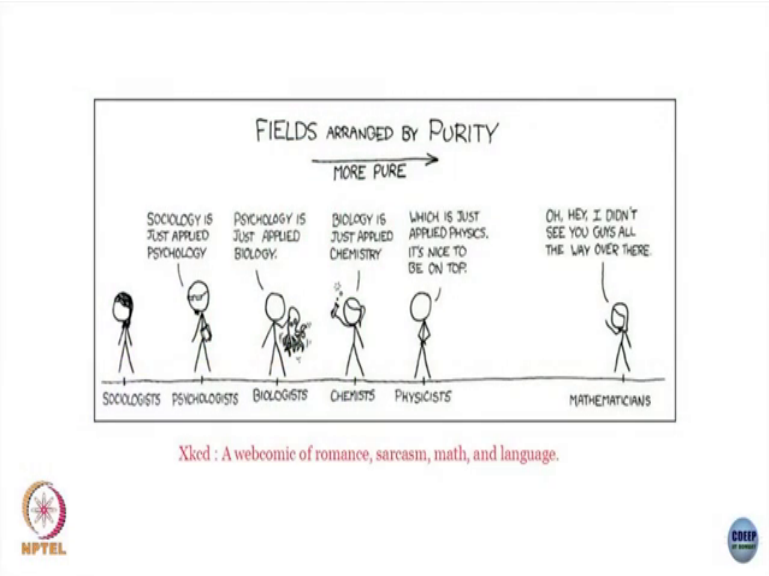
We will try to get an idea of what are the scales and what are the numbers that we are that are going to talk about, how large is a cell, how large is a DNA. Just so, we have an idea of what sort of things at what scales we are trying to model stuff, ok. It is good to have a physical intuition about the scales of things that we are trying to model so, that you know what sort of modeling methods are going to be appropriate.

(Refer Slide Time: 01:07)



So, that is what we will do for the first couple of lectures and then on third lecture onwards hopefully we will start a little bit of actual formal modeling, alright; so, let us start.

(Refer Slide Time: 01:17)



So, before I start I always show the slide. How many of you read xkcd? Ok. So, this is a fields arranged by purity where there sociologists, psychologists, biologists, chemists and physicists and a psychologist says that sociology is applied psychology. Biologists says that psychology is applied biology and a chemist says that biology is nothing, but applied chemistry. And, physicists as most of us are looked down on all of this and say that well everything is after all applied physics and then you have the mathematicians over there all the way, they live on a different planet. So, they do not come into the picture at alright.

So, this is in some sense a very how do I say old school way of thinking where you think how fundamental is my subject in some sense you know. Mathematicians deal with the most purest, most abstract stuff out there; physicists we like to think that you know we are dealing with the fundamental forces, the fundamental building blocks of nature and so on and so forth. So, as you go down on the scale we think that you know you are getting more and more

into non-impure stuff, maybe you look down a little bit on psychologists and sociologists and so on.

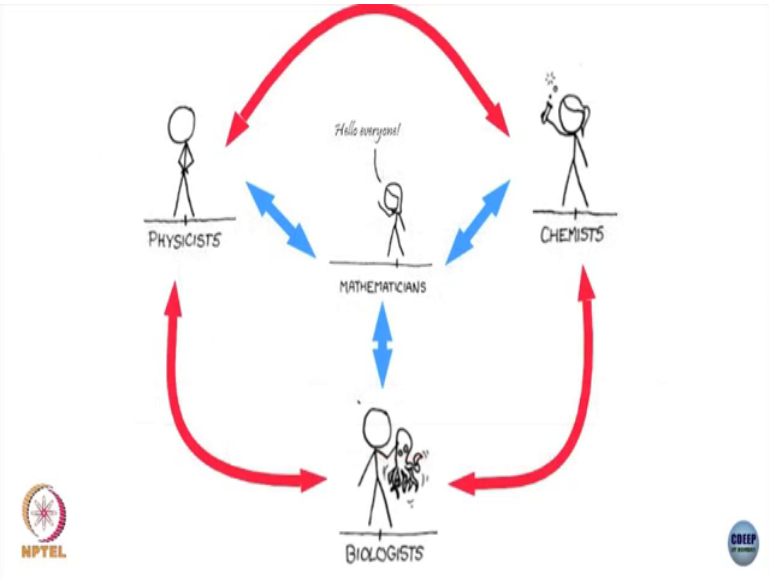
But another way of sort of thinking about this is that as you go down the scale from physicists down to chemists and down to biologists and so on; what you are doing is that you are dealing with more and more complicated systems. Mathematics in some sense is one of the easiest things to study, because you can define exactly the problem that you want. It can have no relevance at all to any real world thing, but still you can define it and you can study it and so on. Similarly, in physics you might have heard of the spherical cow joke, if we like to make as many approximations as we can in order to do what mathematics we can, right.

It may not ultimately map exactly on to the problem that we are studying, but we do not let that stop us, right. We do as much approximations as we think is right, but as you go into biology and as you go into psychology and so on. These are extremely complicated subjects which depend very in crucially on the interconnectedness of various parts. You cannot often study one part of the system in isolation and say that let me try to understand this, let me try to forget about it everything else ok; that strategy which often works in physics or in mathematics often does not work in biology.

We have to look at the whole system, you have to look at how different parts influence different other parts and then try to gain a holistic understanding. Does that mean we do not make approximations? Of course, not we do otherwise as physicists we will be out of business, but we try to keep in mind that what answers we are getting may not often correspond to what experimentalists are measuring.

So, these are really really complicated things and of course, psychology, sociology where you study you know societies as a whole, those are immeasurably more complex. We really have no handle on how to mathematically study these things to any degree of accuracy.

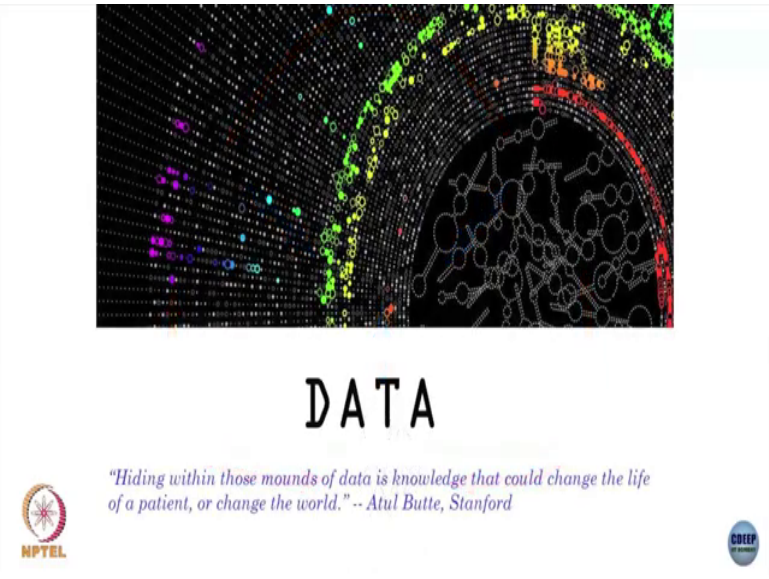
(Refer Slide Time: 04:25)



So, the way we like to think of it now the buzzword of course, is like interdisciplinary right. This whole top, this whole course's sort of interdisciplinary and the idea is that we all talk to each other. I talked to a chemical engineer, I talked to an electrical engineer, I talked to a computer scientist, I talked to a mathematician and all of us try to come together and try to solve maybe one problem in biology, ok.

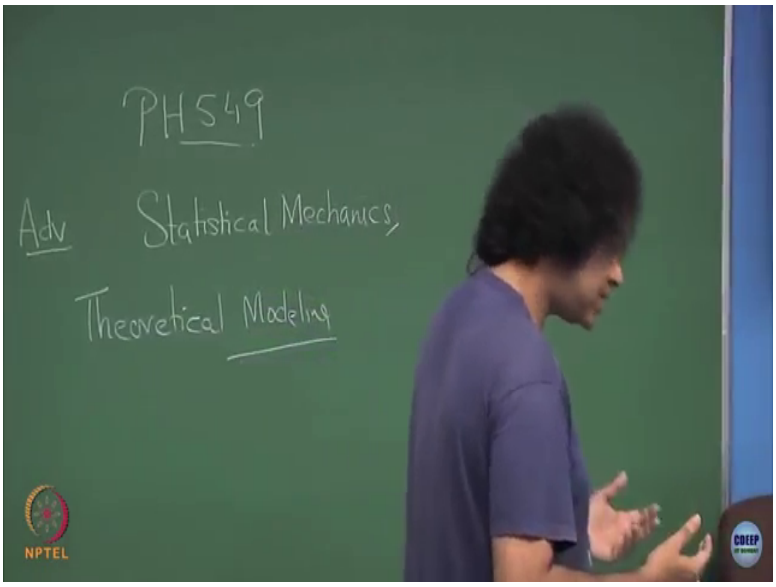
The things that we learn in physics, the things that we learn in computer science, the things that we learn in chemistry all of them can we need inputs from all of these fields in order to gain a comprehensive understanding of maybe some problem in biology, ok. So, that is the spirit in which we will try to approach this course, we will take different things from different places and then try to build models for biological systems.

(Refer Slide Time: 05:11)



And, what makes this course possible although I will not this is going to be mostly a theoretical biology course. So, I will mostly be doing theoretical modeling, right.

(Refer Slide Time: 05:27)

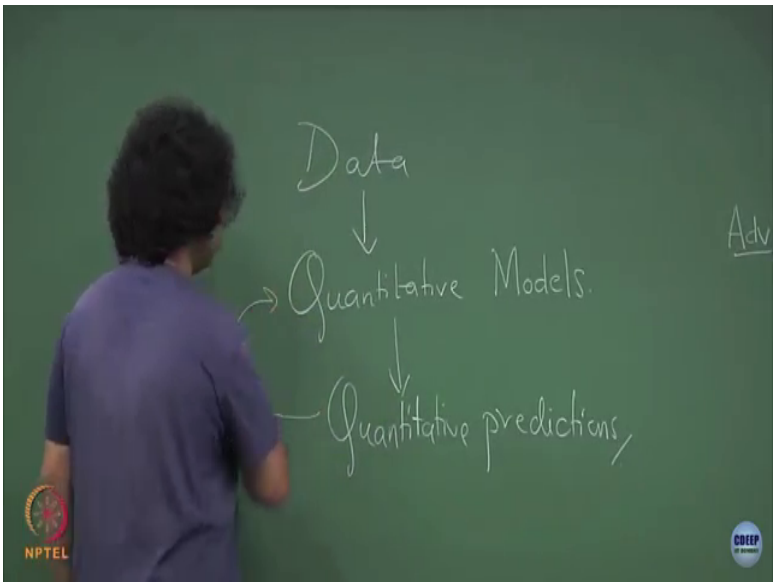


But, it is very very important to keep in mind that this whole topic of physics or biological systems or mathematical biology whatever you want to call it, that it sort of driven by this explosion of data that we have had over the last 20-30 years. When we were in school at least we used to think of biology as somewhat of a descriptive subject you know your DNA and DNA mean proteins. And, these proteins did stuff and that stuff allowed us to live, you had chemical reactions going on and so on. So, it was all very descriptive, but over the last 20-30 years brilliant scientists have made immeasurable progress in devising new experimental techniques that have yielded a wealth of quantitative data.

When I say that a gene is activated, how much is it activated, can I measure it? If I say this much mRNA is produced, this much mRNA produces this much proteins, all of these questions can now be answered quantitatively. So, they can do very high quality measurements which gives us very accurate data numbers basically and these gigabytes and

gigabytes of data have being generated by these very high quality experiments. So, once you have this sort of big data, this sort of wealth of data what you require of what you when you have this data what you require a sort of quantitative modeling, ok.

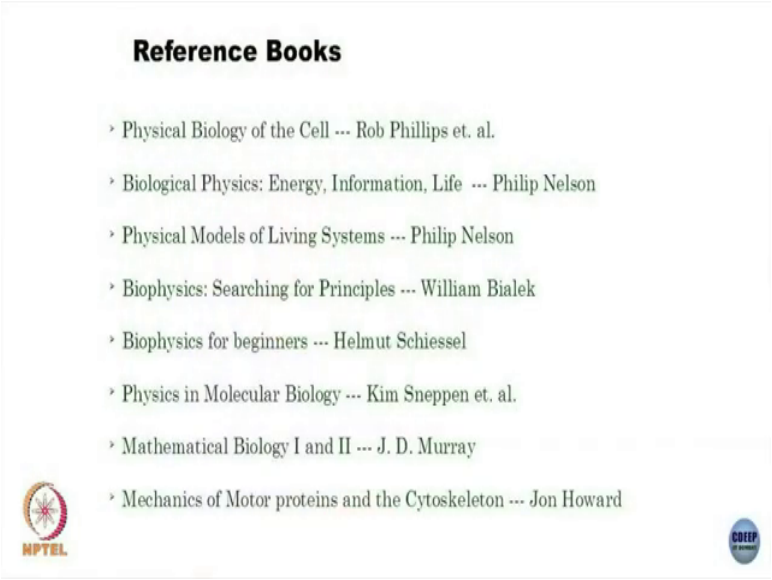
(Refer Slide Time: 06:45)



You can no longer say that well I understand DNA produces proteins and that is it, but can you build a model that will predict how much protein is going to be produced at this moment of time or any future instants of time. See what you want to do is you want to build quantitative models and you should use these quantitative models in order to make quantitative predictions right, that is the whole point quantitative predictions, ok. And, then depending on how good your predictions are you go back a new sort of rebuild your models and so on.

So, it is a iterative process until you get an explanation of the biological phenomenon that you are happy with at a quantitative level, ok. So, that is so, again I will try to speak about the experiments as and when I can, unfortunately I am not an expert on many of these experiments or on any of these experiments rather to be fair. So, I will give I will be able to give only broad hand waving explanations, but still if you are interested you should go back and look up some of these very fascinating experiments as and when we come across them in the course. It is these experiments it is sort of make this whole field possible, ok.

(Refer Slide Time: 08:21)



So, before I progress here are some books. So, one of the problems in teaching physics or biological systems course is that there is no textbook as such, ok. So, that is something that you will have to deal with; I teach I take stuff from some of these books various stuff from various books, various other stuff from papers which I will refer to as and when I use them.

But, if there is one book that you need to study respective of whether you do this course or not, if you are interested in sort of understanding a little a bit about how modeling in biology works how this sort of quantitative modeling works; I would urge you to read this first book over there Rob Phillips Physical Biology of the Cell.

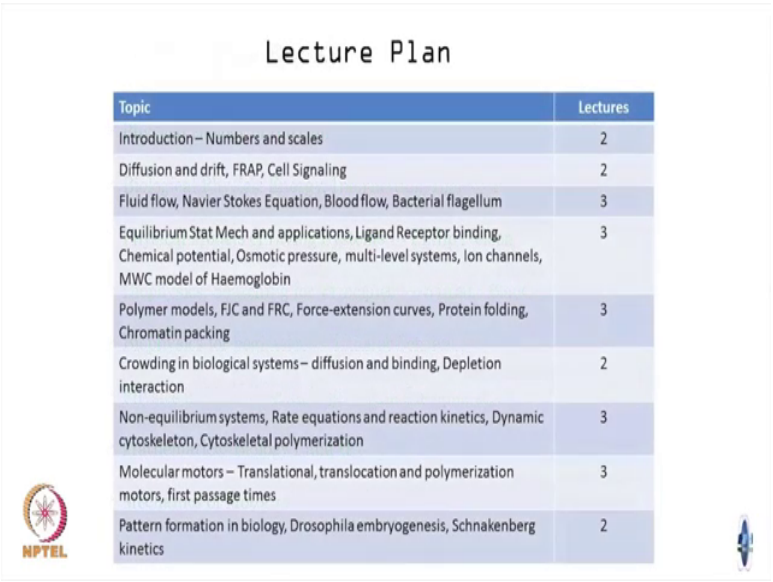
It is an extremely thick book, over this whole course maybe we will cover 40 percent, 50 percent of the topics that is covered in Rob Phillips. The rest I will not even attempt to touch, but it is a very nice book. It is deals with many many different systems, how, what sort of models are there in the literature, how you should think about these biological systems with a plethora of examples. Next these to the Philip Nelson books are also very nice; they are not as comprehensive as the Rob Phillips book, but at least the systems that they talk about its very nice for a physicist or for a very basically a mathematical biologists. It will be read about how to go about building models of biological systems or living systems.

I will again refer to some of both of these books from time to time. The next book is I will not refer to at all Will Bialek Biophysics: Searching for Principles, but I would urge if any of you are interested in the topic to read. So, it is sort of so one of the things that makes teaching biophysics a little difficult is that you know there is it is not a very well comprehensive linear theory. You do not go, not study topic a and then topic b and topic c and topic d; if 5 people were to teach you they would teach you 5 completely different things, that is because we do not really have a very fundamental understanding of what makes life tick, ok.

We understand bits and pieces, we understand how genes work, how DNA produces proteins and so on. But, is there any overarching principle that is not yet clear and that something that Bialek's book he argues for his own pet theories of course, but its provocative to read the book. If you are interested the Searching for Principles as to how should I think about biology in a comprehensive way ok, how what sort of approach like I take to thinking about biophysical problems. These two books I will maybe when I do a little bit about molecular motors I will refer to the next two.

Mathematical Biology by Murray is will completely to us end of the course when I talk about patent formation and this Jon Howard's book is a very nice book on Motor proteins which I will deal with again. I think I will come to motor proteins post medicine ok. So, there is not going to be any one book, it is going to be a combination of different topics from different books, but still if you are looking for any one book I would urge you to read Rob Phillips, alright. So, that is the idea I will of course, upload these notes; so, there is no need to write down all of this, you can just check out more, alright.

(Refer Slide Time: 11:51)



So, here is the lecture plan, I just want to show you this so, that you are not under any illusions as to what we will or will not study. Biology is like I said extremely complicated, you can approach biophysics in different ways. What I will do is I will take the physical approach what is called physical biology, in that I look at how physics forces or physics concepts can be used to study biological problems. So, I will talk about energy, I will talk

about entropy, I will talk about forces and so on. This is not the only way that you can look at biology; you can for example, look at it from an information theoretic point of view.

So, you can look at things like bioinformatics, you can talk about informational entropy and so on which is not something I will cover in the course at all, ok. Or you could look at systems level biology which is again something not I will that I will not cover too much in the course. So, I will only do I will limit myself to a subset of how to approach biophysical problems which is a physical biology approach.

And, even then I will only look at a subset of topics which I hope will be illustrative. So, if you look at this list maybe many of your favourite topics will not be on the list, that is partially my shortcomings. I will teach what I know, also I will teach what I find interesting personally and you have to suffer through it, ok.

So, I will start the first couple of lectures as I said is going to be introduction, then I will talk about diffusion problems, how diffusion is important in biology. I will talk a little bit about experiments such as this FRAP which is a fluorescence recovery experiments. I will talk about the cell signaling process; from there on I will move to fluid flows which is very important because the environment the biological environment, the environment in which a cell lives is very different from the environment in which we live. So, one needs to understand low Reynolds number words in order to understand how biological stuff moves.

So, that is what I will do for the next couple of lectures and then I will go to equilibrium statistical mechanics and its applications and their various applications. So, I will pick up a few examples where you can use the concepts, energy, entropy free energy and so on and see what that predicts of biological systems. I will move, then I will move on to polymer models. So, the DNA is for example, is the polymer, proteins are polymers and so on. So, look a little bit about how polymer physics, how to describe polymers from a mathematical point of view. I will talk a little bit about protein folding and little bit about chromatin packing that will sort of take care until the mid-sem hopefully; post mid sem we will look at crowding.

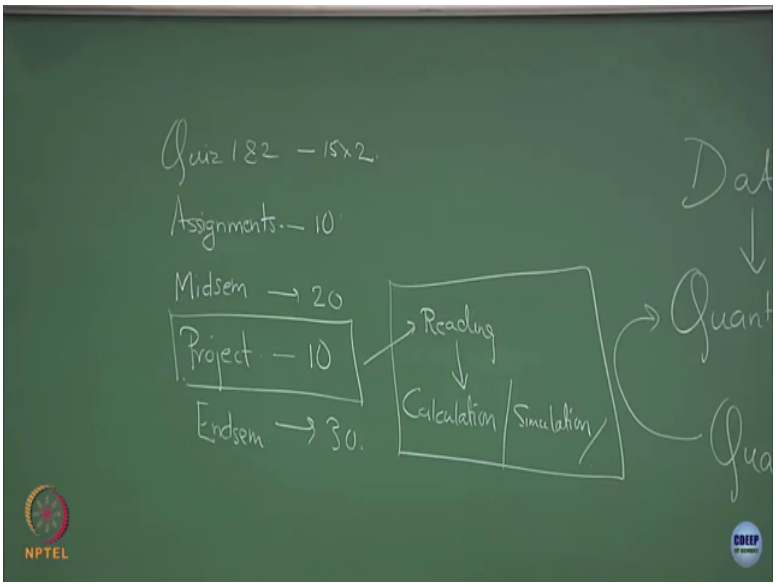
So, pre mid-sem it will be sort of ideal biology in that we will make various approximations, post mid-sem I will try to relax those approximations. So, one of them is dilution approximation. So, when I talk about ideal gases and so on, you know the ideal gas law is valid PV equal to nRT is valid for dilute gases right, when things do not see each other. Similarly, for a lot of these pre mix and stuff I will talk about things in the dilute limit where proteins do not see each other, but different constituents do not see each other. So, one way to relax that or one way that is wrong is that biological systems are extremely crowded.

And, I will talk about what are the implications of that crowding on diffusion, on binding will basically on things if you have done pre mid-sem. I will talk a little bit about non-equilibrium systems and how to deal with non-equilibrium systems in a biological context both for cytoskeletal. And, again if you are not familiar with these terms hopefully this class, in the next and so on will become familiar, but you please stop me and ask if you.

So, molecular motors and cytoskeleton are examples of non-equilibrium systems and I will try to show you how to sort of write down the equations that govern these systems and how to make some predictions for them. And finally, I will end with patent formation, how patents form in biology which are basically examples of reaction diffusion systems. And, then I will apply that to or in very model way I will try to apply that to do relevant systems such as drosophila embryo and so on. So, that is the basic plan, it is a very limited it is again a very very limited subset of what plausibly could be covered in a biophysics course.

Generally, what happens is that after I have finish this I have 2-3 lectures depending on what is the holiday schedule for this semester free. So, in those 2-3 lectures you are free to suggest, if you want me to talk cover any special topic I will try my best to do that, alright. Also before I proceed let us just get a little bit of logistics out of the way. So, generally I have two quizzes. So, quiz 1 and 2, I will have assignments which I will upload on Moodle, there will be a mid-sem.

(Refer Slide Time: 17:21)



Because, this is an elective course and I assume that all of you are here or whoever will stay will continue because you are interested in biophysics. There is also a project component, there is also a project component and finally, there is the end-sem, ok. So, let us say 10 marks for projects, 10 marks for assignments, 15 into 2 for the mid-sems that is what 30, 40, 50 (Refer Time: 18:12). Does that work? 30, 40, 56; 30, 40, 60, 70, 100 yes ok. So, that is going to be the basic break up you will have two quizzes. So, I will upload assignments. So, if that is roughly the number then each of you can do an individual project.

So, you can pick up any topic that you are interested in, since I assume you are interested in biophysics is the topic hopefully there are some topics in biophysics that interest you. So, you pick up a topic that you would want to do. So, a project can be either a reading project, it can be a reading project where you read a couple of papers on the topic. But, ideally I would also

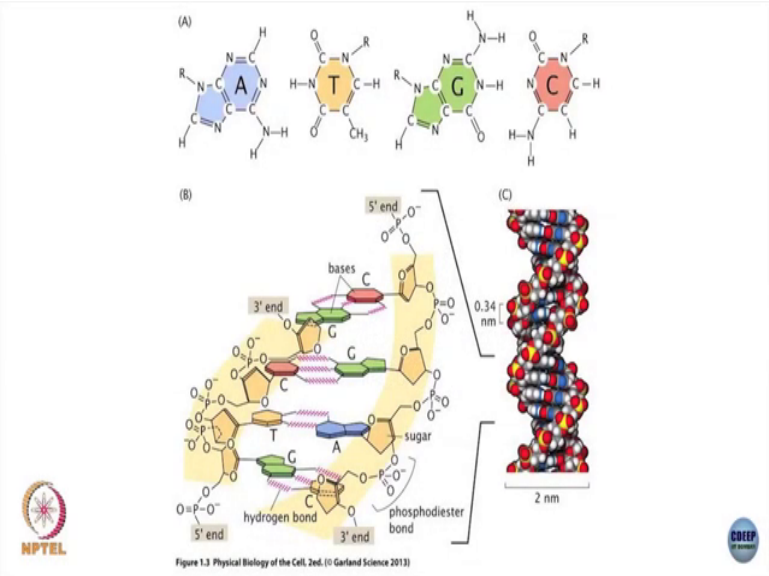
like you to do some calculations either from the papers or if you want to extend any small extension.

So, ideally I would also like you to do some calculation or some simulation ok. So, together and then towards the end of the semester we will find a slot where you can present whatever you have done throughout the semester. So, that is why I want to sort of fix the topics a little early; so, that you get some time to read some time to code or some time to do some calculations.

So, after the first couple of weeks are over, please come and tell me what topic you would like to do your project on. If you have a broad idea of the topic, but you are not sure of where you should start from, what papers you should start from you are welcome to drop in to my office, I am happy to help. You at least find out some introductory papers that you can start reading off with. But, at least in 2-3 weeks you should I should have a complete list of who is doing what project, ok. And, then towards the end of the semester you can present whatever you have done clear, good.

The assignments of course, I leave up to your own integrity, try to do them on your own. I have no way of ensuring that, but I hope you know because, again because this is an elective course and hopefully you are doing it because you are interested. Please try to solve the assignments on your own or at least understand if you are copying, ok.

(Refer Slide Time: 20:41)



So, introduction right; so, let us start with the basic sort of building block of biology which is DNA, right. And what is DNA made up of? DNA is made up of these 4 nucleotides which are over here; adenine, thymine, cytosine and guanine. Adenine, thymine, guanine and cytosine, ATG and C, these make up the 4 alphabets that make up a genetic code. So, I have these 4 nucleotides and one picture of DNA is that it is nothing, but a repeating string of these alphabets ATG and C. So, it is a very simple language in that sense there are only 4 alphabets and all the information about, it is just scares me sometimes too.

All the information that is contained in our genetic code is contained in these alphabets of ATG's and C's, ok. These are arranged as you know in a double helix, the famous DNA double helix and this double helix sort of encodes all, not all, ok. This double helix sort of encodes the information that makes us what we are, that makes us alive, that makes each individual species what they are, ok. And just to get a sense of the scales each so, these are

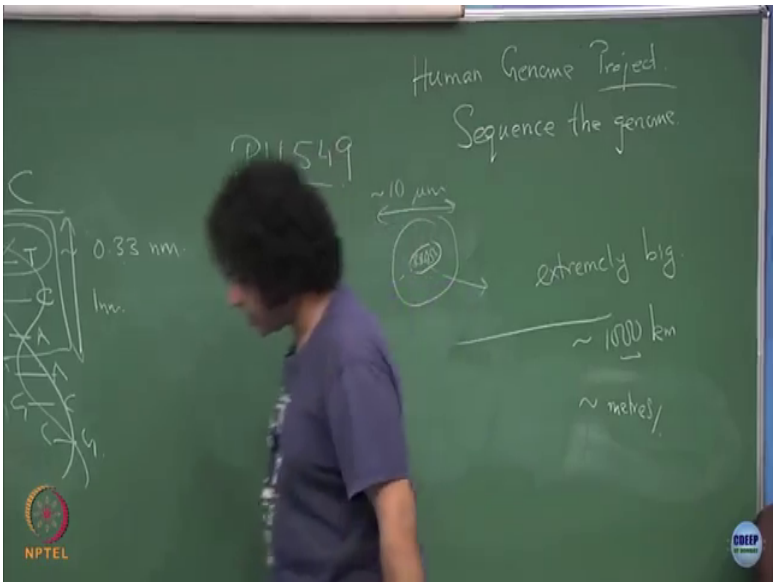
called base pairs; so, these are called base pairs. So, when you have a double helix like this A G T T G C and so on and adenine always pairs to a thymine on the complementary base or guanine always pairs to a cytosine.

So, like that G C C G, right. So, you have these two strands and you always have these complementary base pairs under on the two strands, right. So, each base pair the size of each base pair is typically around 0.33 nanometers or what it says here is 0.34 nanometers; so, one-third of a nanometer roughly, ok. So, if you took three of these base pairs like this together this thing would maybe make about 1 nanometer. So, that is the scale that we are starting off from and the width of this thing over here is around 2 nanometers, ok. How big would the DNA let us say let us talk about a particular DNA, let us talk about the human DNA. How big would the human DNA be, if I stretched it out any idea?

So, if I took this you if I took the DNA that is present in ourselves which is inside. So, here is my cell, inside my cell is my nucleus, inside my nucleus is all my DNA right. If I took this DNA and I took it out and I stretched it out how long would that be; any idea? Anyone?

Student: Extremely big.

(Refer Slide Time: 23:55)



Extremely big, there is no way that answer can be wrong, right. So, correct answer, but so extremely big. How big is extremely big?

Student: Thousands of kilometers.

Thousands of kilometers; thousand kilometers, anyone else? Anyone else wants to take a wild guess as to what is extremely big? How just to get a sense; so, we know that each base is one-third of a nanometer, right. So, here in this DNA are my base pairs with each base is one-third of a nanometer. How big would cells typically be? Order of order of magnitude, micrometers; so, this would be of the order of 10s of micrometers maybe, right. So, 10 to the power of minus 6 meters and therefore, the nucleus is of course, proportionately smaller. So

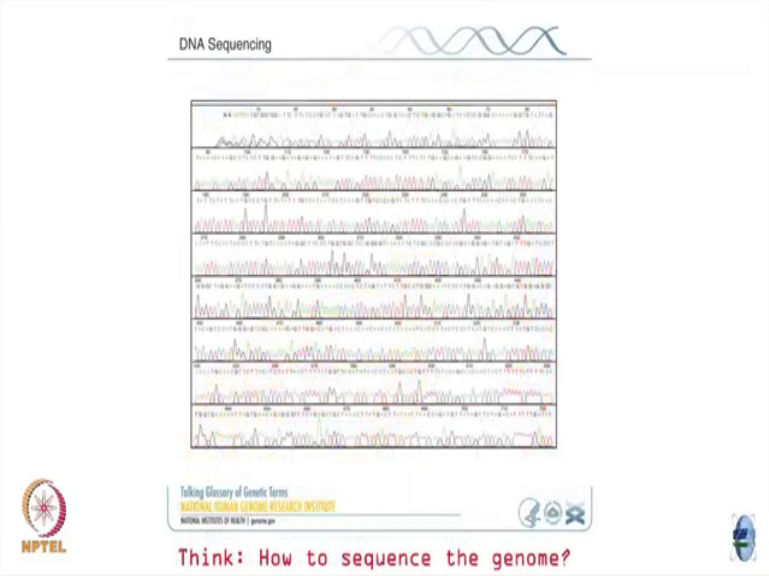
however, long that object is; however, long that extremely big object is you will ultimately have to pack it into something which is of the order of microns, ok.

So, that is of course, a challenge that does not mean that that is wrong of course, the DNA is extremely big; the question is how big is extremely big. Anybody else? No.

Student: Meters.

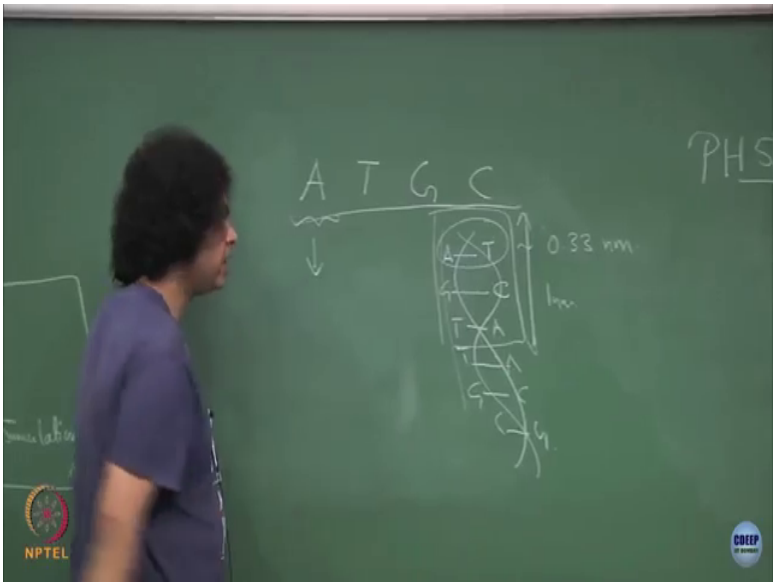
Meters; so, this is of the order of kilometers, this is of the order meters; I think meter is roughly correct. I have at least 4 human DNA, some 3 meters or something. So, you take that 3 meter long chain which is made up of these 0.33 nanometer building blocks, the base pairs and then you package it into the chrome, packaged it into the nucleus.

(Refer Slide Time: 25:43)



So, as you know over the 90s and so on, starting from the 80s and then in the 90s with the human genome project, one of the central challenges of this sort of quantitative approach to biology was to determine the sequence of the human genome or the sequence of any other animal genome or any other living species genome. So, one of the central challenges was to sequence the genome which is to say that the idea was that all the information, all the genetic information that I am passing on through reproduction that is contained in the sequence, in that is in your DNA.

(Refer Slide Time: 26:23)

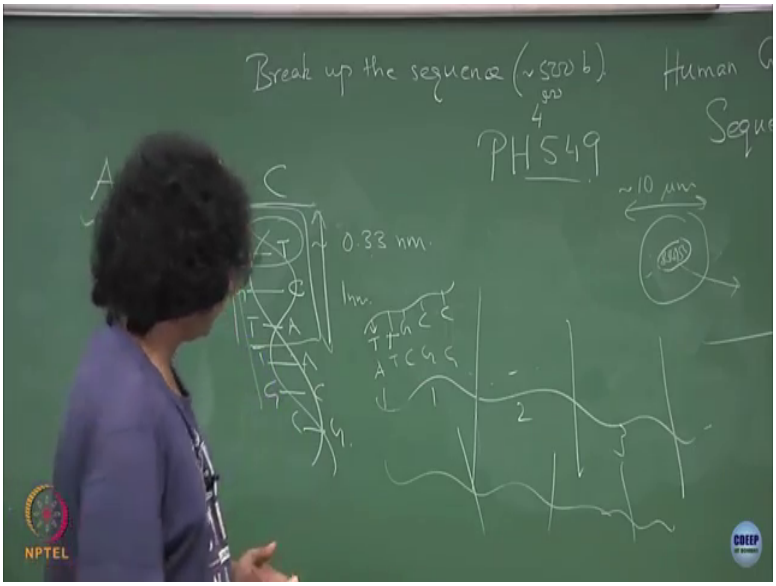


So, can I understand what the sequence? Can I know what the sequence is? Right, A G T T G and so on so, for millions and millions of base pairs, can I know this complete sequence? So, that was the challenge of the human genome project which is now which was successfully done. So, it was sort of marvel of science and a marvel of technology that we managed to sort

of read out exactly what is the sequence of the genome of a species including a human species, ok.

So, any idea how if you were given with a sort of string so, basically my I can think of my DNA as a string, right. This string has these alphabets A T G C and whatever millions and millions and billions of these alphabets in a linear chain. So, if that is all the information that I need in order to understand why a cell behaves as it does, I want to under I want to first read off this whole sequence of alphabets.

(Refer Slide Time: 27:31)



How would you go about doing that? Any guesses or any sort of random ideas?

Student: Cut it.

Cut it, ok. So, I cut it then?

Student: Do that to multiple DNA.

Do that to multiple DNA strands. So, I take multiple DNA strands, I cut them that is that is right, that is the first step to actually reading the sequence of a genome is to actually cleave the genome into very very small tiny bits, ok. And, then people have found out fancy ways how we can reconstruct them back together. So, if this is strand 1 and this is 2 and this is 3, how do you know where 2 comes and 3 comes and so on. So, that is part 1 which is you break up the genome.

So, break up the sequence; break up the sequence, what then? How do I know which one is A and which one is T and which one is G and so on? Ok. So, you make a complementary strand is that what you are saying so, ok. So, that is a nice idea, I do not know if people have tried it out. The question is so, let us say let me design the perfect complementary strand right, T A G C C, right. This would go and stick to this and then you would say that well since I know the sequence of this complementary strand, I can tell what is the sequence of the original strand maybe I just do not know how sensitive it would be given that so, ok.

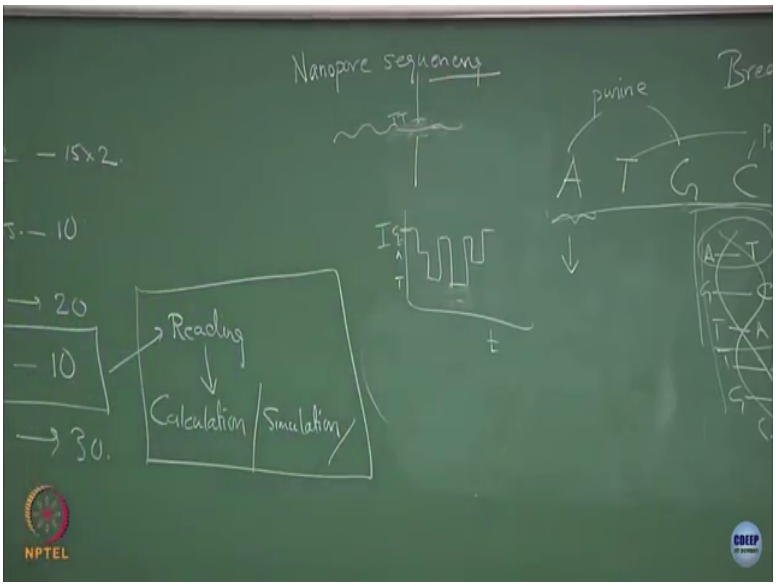
So, one thing to know is that how big are these strands they typically cut into they are not very big, they are not 5. You typically as far as I know I do not know exactly you can look it up; this is about 500 bases, ok. So, in each position you can have 4 letters, 4 alphabets. So, you have 4 to the power of 500 possible strands that you need to generate. Secondly, the question is how specific would you be? So, if I were to make this one a wrong thing right and everything else correct, this would still go and stick together probably, right.

So, I am not; the question is whether you could quantify the stickiness? You can look it up, I am not an expert on this and there are various ways actually, but this is an interesting way. So, you design a complementary strand and then try to see whether you can map this original strand to the complementary strand know ok.

So, then you can take this as assignment 1, on Friday you can give me a 1 page write up as to how to sequence the genome ok, people do different things. So, you can look up what things people do, you can choose your favorite method and you can tell me how they do it; so, that I also get to know basically.

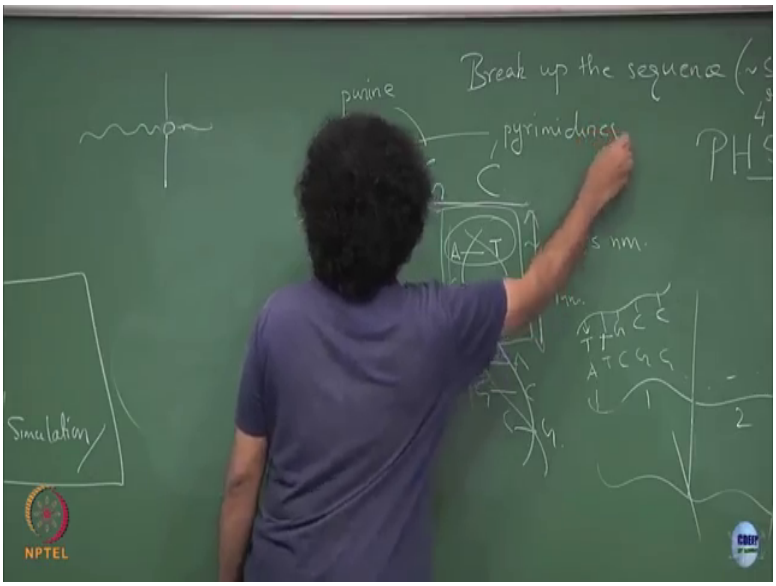
So, this is your assignment for Friday ok, 1 page A4; no need to basically write the central ideas to how people sequence the genomes. I know one way which I can tell you, I do not think that is the way people do commercially, but that is something that some of my collaborators used to do; so, I can tell you. What they do is that they take a pore ok, they take a very very small nanopore basically or it could be a protein pore. So, this is of the order of nanometers and you thread a DNA strand through this pore, ok.

(Refer Slide Time: 31:19)



And there are elect; so, there is let us say some current passing from this side to that side and you read off what current passes. So, when so each base, each of these molecules A T G and C are of different sizes, right. So, each of them are of different sizes. So, when each of them when a T passes through a let us say A and Gs, if I remember correctly are the big ones, these are called the purines.

(Refer Slide Time: 31:47)



These are the ones that contain two of these As and the Ts and the Cs are the smaller ones, they are called the pyrimidines; I hope I am right, ok. So, each of them are of different sizes. So, you can see that A and G are big, T and C are small. So, when each of them let us say when a big thing blocks this pore, the amount of current that passes through is actually smaller then if a small thing blocks this pore. So, if you have a small thing that blocks this

pore, the current that passes through this from here to here is larger than if a bigger thing were to block this pore.

So, you first send the known strand and you see; so, if I send the strand of Ts for example, you see what is the current reading, ok. Similarly you do it with Gs and Cs and As and then you take the strand that you actually want to sequence and you measured the current as a function of base pair, function of time. So, what you will see is that you will get four levels of the current maybe like this, ok. So, you can say that well this level corresponds to maybe G, this level corresponds to a C, this level corresponds to maybe I have drawn five levels, this level corresponds to an A and this level corresponds to a T.

So, by reading off the current that is passing through the pore you can get readout of the sequence of the DNA strand that you are passing through the pore. This is called nanopore sequencing, this is called nanopore sequencing. I am not, I do not think this is the method that most commercial sequencers now use, they probably use some sort of electrophoresis, but I am not sure so, anyway.

So, you can find out what are the actual sequencing techniques that people are nowadays using, it is become quite cheap to actually sequence genomes. So, it is become very fast and very reliable. So, you can find out and tell me on Friday how the sequencing the genome goes.