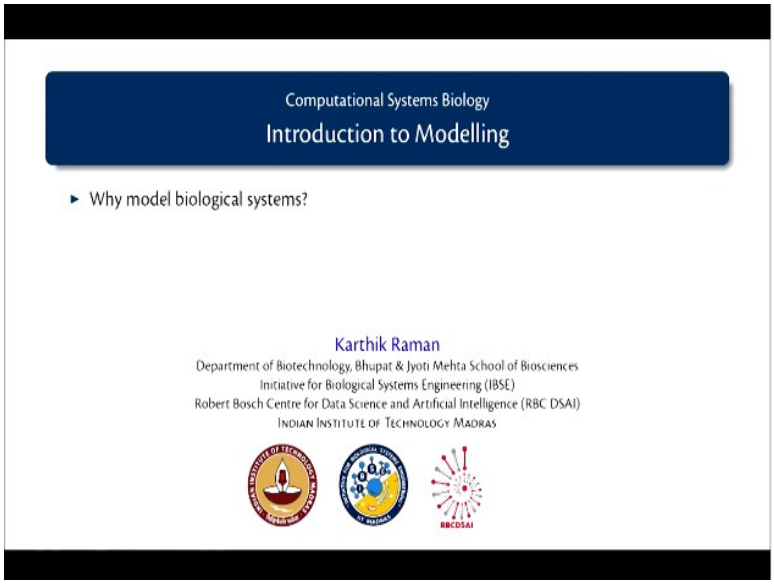
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

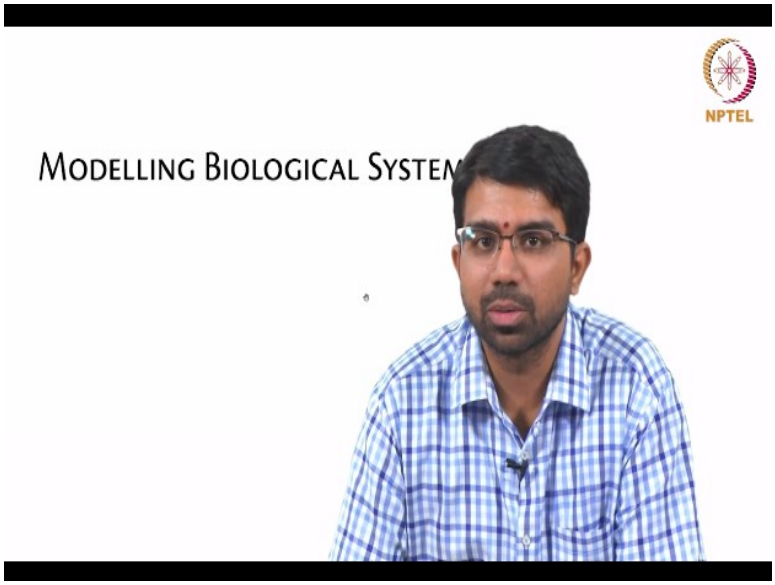
Lecture - 02
Introduction to Modelling

(Refer Slide Time: 00:11)



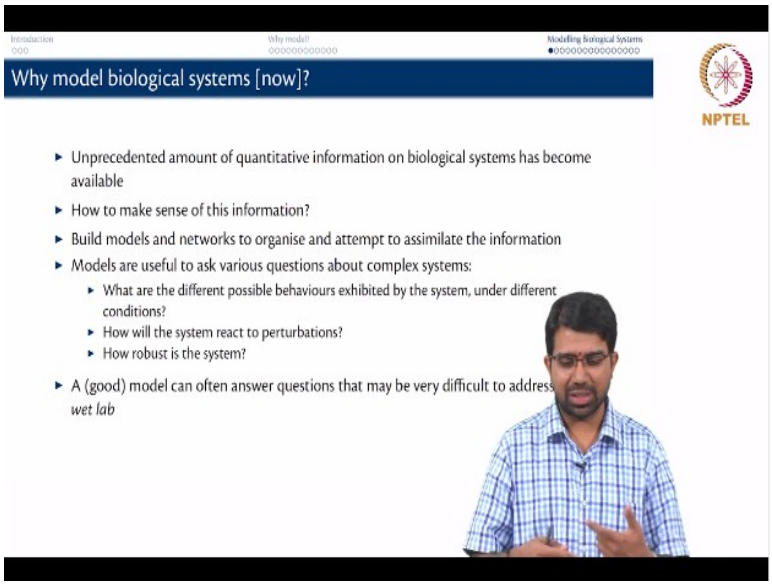
In this video, we will continue with an introduction to modelling and we will focus on why model biological systems? And what are the challenges associated with them?

(Refer Slide Time: 00:22)



Welcome back! We will look at how we go about modelling biological systems. We started with the basic introduction to modelling in the earlier session. And now, we will look at what are the challenges with respect to modelling biological system.

(Refer Slide Time: 00:37)



So, why do we model biological systems now? What has changed? So, if you see as we have all heard about the human genomics and the various omics projects that have been going on in the recent past. And you must know that there is an unprecedented amount of quantitative information that has now become available. Biology and math are sort of orthogonal, right?

And especially in school and all that, you find that you either like math or you know biology. If you don't like biology, you do math. If you don't like math, you do biology. But things are now changed, in the sense that you need to be good at biology and math as I will show you, may be in the rest of the course and also particularly today, as to why there are so many interesting and important computational challenges that need to be addressed to better understand biology itself.

So, that is one important thing to look at. And how do you make sense of all these information? You must have heard of terabytes and terabytes and terabytes of genomic, transcriptomic, proteomic, metabolomic information that are coming in. So how do you make sense of all this information? So, one way to go about it, is to try and build models. Because models as we talked about in the morning can help you make some sense of really complex systems.

So, how do you make sense of complex systems? You try to build models. What kind of models do we build? That is what we will try to look at in the rest of the course. Models can basically answer questions about really complex systems, right? For example, what are the

different kinds of behaviours that the system can exhibit? Question number one. And how will the system react to perturbations? Or how robust is the system?

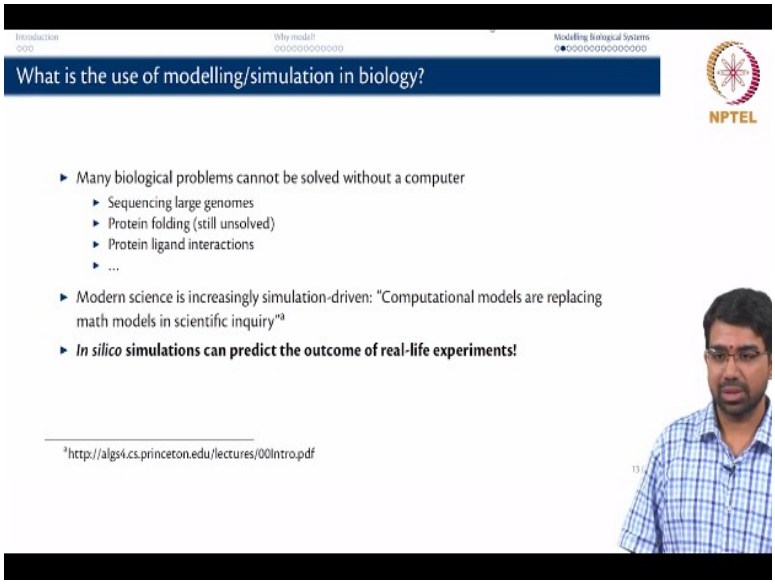
Perturbation is the very important theme as you will see again and again during this course. You will see that perturbations are very important. How do you understand biological systems? You try to perturb them and try to see what happens. You remove something, you figure out what is the effect of that. So, if you go back and read the paper on "Can a biologist fix a radio" which I have already uploaded, you will see that this is one thing he talks about.

This is very different from other sciences. No electrical engineer tries to study a system by pulling out a component and figuring out how the pull out affects the system, right? But in biology, you are pressed for choice. There is no other way to really study a system nicely. This is one window and as one of you was mentioning in the morning, you will see that this "what if" question of you remove something, what happens when you remove that, right?

This question becomes very important in trying to understand biological systems and similarly when you model biological systems. And a good model can often answer questions that will take a few PhDs in a wet lab. So, a good model and note the operator word there, is "good". And what do we mean by a good model? It's something that works across different conditions, probably has fewer assumptions, right? And has probably been validated, most importantly.

These are all the things that we will look at as we look at what the modelling process is and so on.

(Refer Slide Time: 03:50)



So, what is the use of modelling or simulation in biology? Do you know of many problems that need modelling and simulation in biology? There are a few here but in fact these are all actually not modelling problems, they are more computational problems. But what are the modelling challenges? See, we did discuss earlier, right? So, why do we model in the first place? Even before understanding? Understanding is the loftier goal.

What is the need? Prediction! Prediction, prediction, prediction. So, that is why we have climate models and that is why we have all kinds of models, right? Prediction is a very key aspect of modelling. What is it that you want to predict in biology? Which is the other way of saying what is it that you want to model in biology? What are the challenging modelling simulation problems in biology?

So, how do you synthesize a metabolite? This is a little more advanced but you could start. Fundamentally, what is life? Life has at its root, DNA. And DNA is the underpinning of life so to say. And what do you see? You see people, you see phenotypes, right? So, the holy grail of modelling is basically genotype to phenotype. Given a genotype, can you model the phenotype?

We will not do this entire genotype to phenotype modelling but we will bridge several things here. We will look at a simpler phenotype and a simpler genotype. One way of looking at it is, I have this metabolic network. Can you tell me how will it grow? So, now the phenotype has been simplified to growth and the genotype has been sort of amplified to metabolic network, right?

Technically, I should say, if this is my DNA, what will be the corresponding metabolic network for that and how will it grow. So, you basically bridge. This is the level that we usually look at. I have this kind of a network. Something like the social network of the cell, right? This is the social network inside the cell. These are the proteins that interact with one other within the cell. How does it behave?

What kind of functions does it orchestrate? So these are some of the questions that we would like to ask. And if you look at computational problems, the most classic computational problem is sequencing a genome. Some of this are still unsolved. You still can't sequence some really big algal or fungal genomes and so on wherein the genome is so big that you cannot assemble it.

Human genome is already 3 gigabases, right? But there are other genomes that are even larger which you can't easily assemble. Protein folding, of course! I will try to give you a peek of how modelling can be used to solve problems in protein folding. And protein ligand interactions. This is at the heart of every single network that we will talk about.

So, every single network has at its heart, some sort of molecular recognition. Molecule A binds only to molecule B. It doesn't have anything to do with a molecule C. So, it could bind to a spectrum of molecules, maybe it binds A, B, C but it will never ever bind to D. This specificity and the networks that arise out of these are what make the cells so complex and so interesting and give you all the different spectra of behaviours that you can see.

The cell can take different behaviours. See, we can survive in different temperatures for example, right? And that's because we have different ways to compensate for different changes in your environment. And very interestingly, you see that modern science is increasingly simulation driven. And I really like this. Because computational models are replacing mathematical models even.

So this course is essentially about mathematical modelling but I could might as well say that it is about computational modelling, right? The distinction between these two is blurrier and blurrier and we are like raising the level up, we are bringing math models to become very

close to computational models. You can translate math models to computational models with ease.

Let's take a simple example. Have you all heard of molecular dynamics? So, molecular dynamics is not something we will do in this course but molecular dynamics is essentially the science of studying how a molecule behaves. So, you have a particular protein molecule. It is surrounded by a few water molecules. Maybe there is a ligand molecule which can bind to it. What happens?

What will happen to the molecules? They will basically vibrate, they will move around following Newton's laws of motion, right? But if you apply that systematically across say a few tens of small water molecules and this molecule. You try to find out what is the conformation? How the protein has folded? And those kinds of things which is what I am talking about here in terms of protein ligand interactions.

"Professor - student conversation" Yes, it is not just Newton's law. You will have other effects and so on. But it is essentially laws of motion, right? So, the laws of motion will obviously translate differently for such nano scale systems and so on but basically you write some force, F = ma as your equation.
And how do you compute all of that? You have what are known as force fields and really complex mathematical equations. But the thing is, from a basic rule, you have actually built a simulation system.

The rule is that, molecules will follows some laws of motion, whatever. And based on those rules, you have actually synthesized a system. You can add one more component to the system and all you have to do is, the rules basically stay, you just need a more powerful computer to do the simulations because the complexity has increased. Instead of 10 equations you will now have 20 equations. Just the addition of one component will usually double the set of equations that you have to solve and so on.

And many of these *in silico* simulations can actually predict the outcome of real life experiments which is the only reason you will do these simulations, right? If the simulations aren't going to be very predictive they aren't going to be useful. So, many of these simulations can accurately predict real world situations therefore you can use these models as

a substitute right. And there are different kinds of models so maybe we should just stop and think about what kind of models are people using.

And if you start with non-mathematical models, so what is a mouse model? See a mouse model is basically a mouse which is used as a model of human infection. Or whatever, human disease, human whatever. Human phenotype technically, right? So, you assume that I do some perturbations to the mouse. I may be give it a disease or inject it with a carcinogen and I see how it develops cancer.

We believe that this is a model to mimic how a cancer evolves in a human being although this is obviously wrong. It is approximate but it is useful. So, actually these are key aspects of a model. A model is by definition wrong because it only models a part of reality as we saw earlier. And it is approximate as we have very well agreed. It relies on several assumptions which you hope should not be violated.

See it should not be violated in the scope of the model. You can't apply the model to a situation where the assumption stand violated at the start. So, these are some of the things that you need to worry about. But basically you have different kinds of models and to just give you a different picture, let's talk about a non-mathematical model. Mathematical model is what we will do for the next several weeks But what is a non-mathematical model?

So, one nice example is a mouse model. And there are different kinds of models that are commonly used by biologists, even all those biologists who do not like or appreciate mathematical modelling use mouse models. In fact, *C elegans* is a model system. And who really cares about *C elegans*, right? It is a very useful model system to study development, to study something else.

Yeast is a great model system. *E. coli*, best model system ever, right? Best prokaryotic model system. *Saccharomyces* is the best eukaryotic model system. *C elegans* is the best you know slightly higher worm model system and so on. And then, you have *Arabidopsis* for plants, Drosophila for flies and so on. So, incremental levels of complexity. And then maybe at the highest end of that spectrum would be I would say mouse or you can even go higher to primates and so on.

But these are all models which try to help you understand a more complex system. You can use yeast to understand so many difficult challenging questions about say even human biology. And this is something that people need to appreciate and of course there are any number of mathematical models that we will see. So let us just spend some time to think about other models that you may have encountered.

So, do you know of any other mathematical models let's say, that you have encountered in the past? Yeah, so that is the mathematical method like ODEs or PDEs are mathematical methods but what would be the underlying mathematical model? I mean what does it try to model? What is the system that is being captured using those equations? Climate models? Very good example, right?

So, we all know about climate models. They are used to predict whether there will be rain, what will be the average precipitation, all those things. And climate models can be very challenging. You know in some places it is going to be easy to build climate models and in some places it is going to be harder to build climate models. And the biggest challenge potentially in climate models is, the sensitivity to initial conditions is very large, right?

So, the ability to predict therefore rapidly drops. So, knowing today's climate you might be able to predict tomorrow's climate but you can never predict one month down the line right because there are so many possible futures that exist based on these initial conditions, right? So, your distance from the initial point has to be minimal. Chaotic, yeah! I didn't want to use the term because it opens a whole new can of things but that is what chaos is, right?

So, you have catastrophic response to a small change in initial conditions where we use the systems where you change the input by epsilon, the output changes by some function of epsilon. The output is not unrecognizable. But in a chaotic system that is what happens. And biological systems somewhere tread the region between order and chaos. There's obviously a lot of order.

There's a lot of complexity and you could say that they do seem chaotic in several situations. You make a small change in the input, you will have a complex response. And maybe you can just bring it down to the nonlinearity right. Linear system is what we all understand. Linear

system is how we are trained to think in some sense because we see so many lot of linear systems in front of us, right?

When you adjust the shower in the morning it more or less behaves like a linear system. You increase this, the temperature will go up. You decrease this, the temperature will go down. It's not suddenly you increase something, the whole water will freeze but that is what. So, in one direction you go, the response is not going to be quantitatively or qualitatively very different from the other direction. But in a nonlinear system, all kinds of weird things will happen.

So, any other models? Stock market, right? Although, I would wager that nobody understands. They want to build a model for everything instead of accounting for noise. There is always some noise but we always say that the stock price of this company fell by 0.1% because the CEO went to the hospital. So there is always a reason that has to be attributed to everything but we will come to that.

There is this danger of blaming correlation to causation. It is a dangerous road to tread. Fair enough! What else? What other kinds of models? Disease models.. so disease models I would go back to mice and so on. Are you looking at any different disease models? How disease spread within a population, right? That's something we will discuss a little later today itself .

So, how does a disease spread in a population? Very interesting! Epidemiological models. How many beds do I need in the hospital? It is a very fundamental social question that needs to be addressed. If there is a Zika outbreak or a SARS outbreak, what do I need to do? What does the government need to do, right? This becomes a very interesting and important question. **"Professor - student conversation"** Population dynamics, right! Could be in a city or in a place or in a field, in a forest.

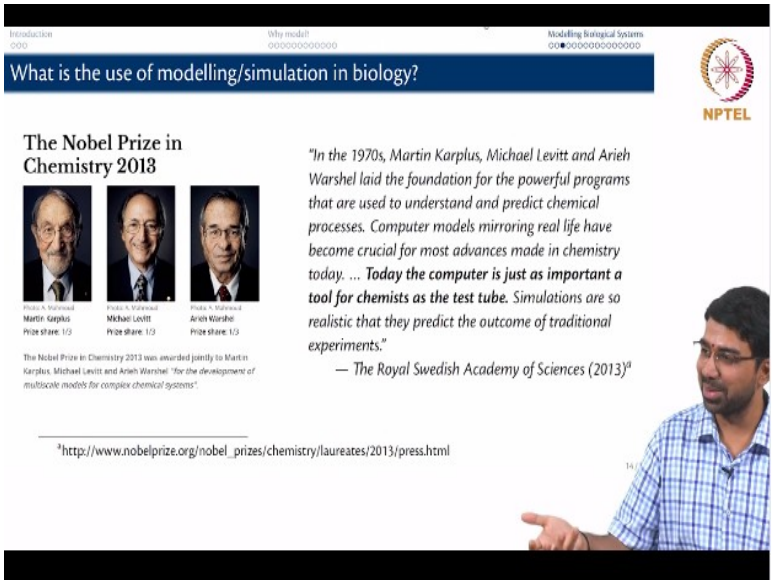
How many lions, how many deers, how many rabbits and so on, right? So, there is a delicate balance and you can always write some sort of equations to say when this goes up, this will go up. When this goes down, this will go down, those kinds of things. Fair enough! Any other models? Neurons, right? So, how do neurons fire in your brain? How do you perceive different things and so on?

You have a whole battery of neural networks and neuronal models and so on. You have any number of mathematical models and you will see that the unifying theme across this is, A. You commit to an understanding of a system, B. you want to really predict something, right? So, you commit to an understanding of the system and by making certain assumptions. And if your assumptions stand violated, you want to go back and fix those assumptions.

You want to see if I am moving to a zone where my assumptions stand violated or my assumptions are fundamentally flawed which means I may have to go in for a different set of modelling tools. I mean no longer we are allowed to use ordinary differential equations, I may have to consider using partial differential equations. And so on and so forth. Or I might have to worry about stochasticity.

I have been assuming that my system is deterministic but may be it is actually stochastic because there are very small numbers of molecules interacting, so that is, now the probability of collisions becomes more important than, average numbers and populations and things like that.

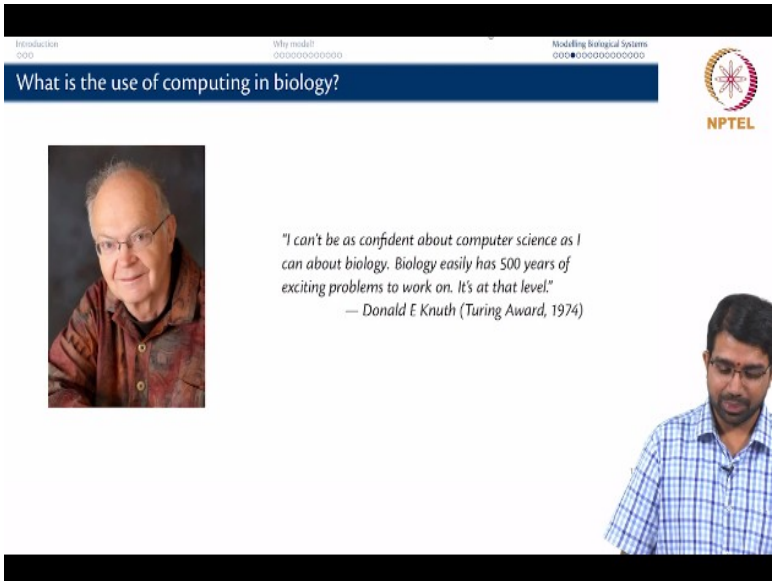
(Refer Slide Time: 18:22)



So, to further drive my point, a very important you know I have just emboldened one sentence from the noble citation which says, "Today the computer is just as important a tool for chemists as the test tube". You could strike out chemist there and replace it with any other you know biologists, biochemist or whatever. Because this goes back to what I was saying in the previous slide in the sense that the computational models are becoming so accurate you can have them replace experiments.

And so you can do more experiments therefore because computational experiments are cheaper to carry out. So, this was a big moment of pride when computational people essentially won chemistry noble.

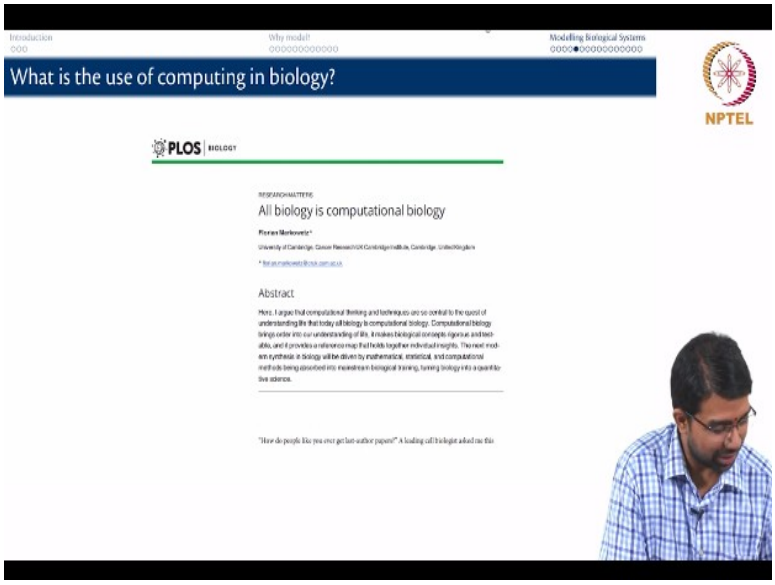
(Refer Slide Time: 19:11)



And of course, Knuth who is easily the greatest computer scientist who already thinks that biology has more number of exciting problems,computational problems although to solve for the next 500 years. And there are so many different kinds of problems and each new technology actually throws up a problem. Nobody worried about read assembly till Illumina came. Nobody worried about read assembly when Sanger sequencing was the standard.

But once Illumina came, you had to worry about read assembly. We will try to look at some of these things a little later on.

(Refer Slide Time: 19:46)

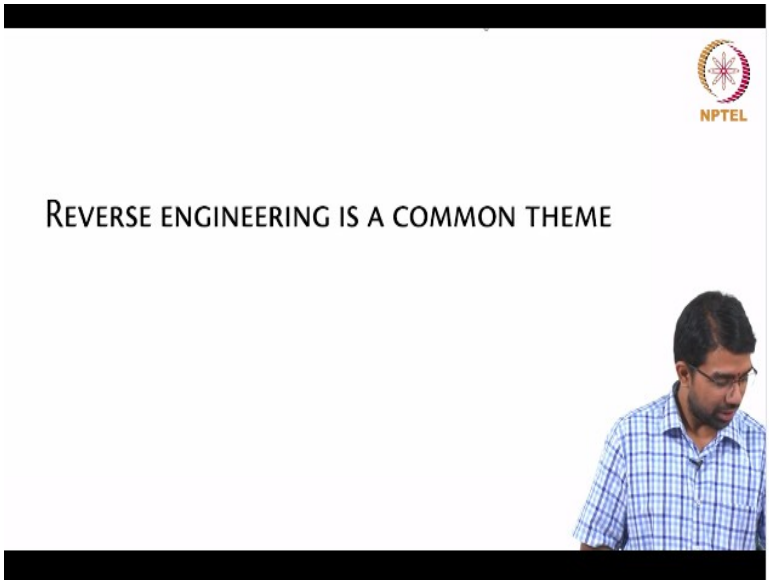


And a very provocative piece that was written earlier this year, which says all biology is computational biology. And this is not new. I think I have been emphasizing it and many people do emphasize and understand that biology and computation are really deeply intertwined, you cannot really separate out the two. You cannot have the lab that does just plain biology or lab that does plain computation.

Well, you can have a lab that does plain computation but they will have to rely on biologists for the data, for the data sets and so on. So, here he basically says that the next modern synthesis in biology will be driven by mathematical, statistical and computational methods being absorbed into mainstream biological training which is what we hope to achieve through courses such as these.

And if you note, I have a small quotation at the end. The initial part of the paper, which is, ‘’how do people like you ever get last author papers?’’ Because bioinformaticians usually play a second fiddle all the time and it becomes a challenge for them to get authorship and all those kinds of challenges start emerging.

(Refer Slide Time: 21:02)



One problem that I was talking to you earlier in the morning today was biology requires reverse engineering.

(Refer Slide Time: 21:09)

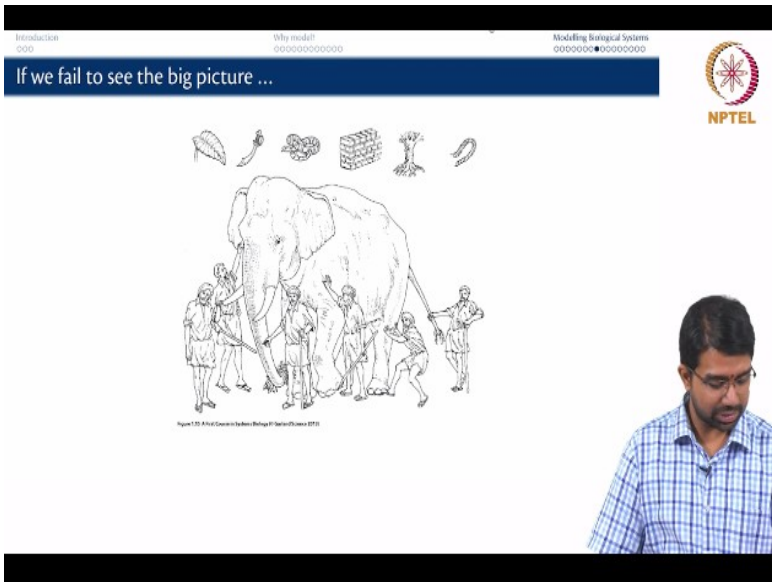
So, can you figure out what this does? I can give you a non-clue which is this is from the US navy Laboratory. **"Student answers".** You can't even guess it, right? But its really tricky because this is how you have to understand a biological system. You start off with the really complex system such as this and then you have to dissect and figure out how the system really works. How does the system work?

May be one thing you can do is, you pull out one of those you know mirrors or microscope looking things there, take it out and see what happens, right? Does it change some output? It will be some output coming somewhere and some input going in somewhere and you mess

up with the wiring and see what happens. But this becomes very difficult, right? And what if you didn't even have the right tools to manipulate the system or you have a limited set of tools that can affect only a part of a system.

These become the challenges and this by the way happens to be that cesium clock. The atomic clock which records several vibrations of the cesium atom and this is the standard by which time is defined.

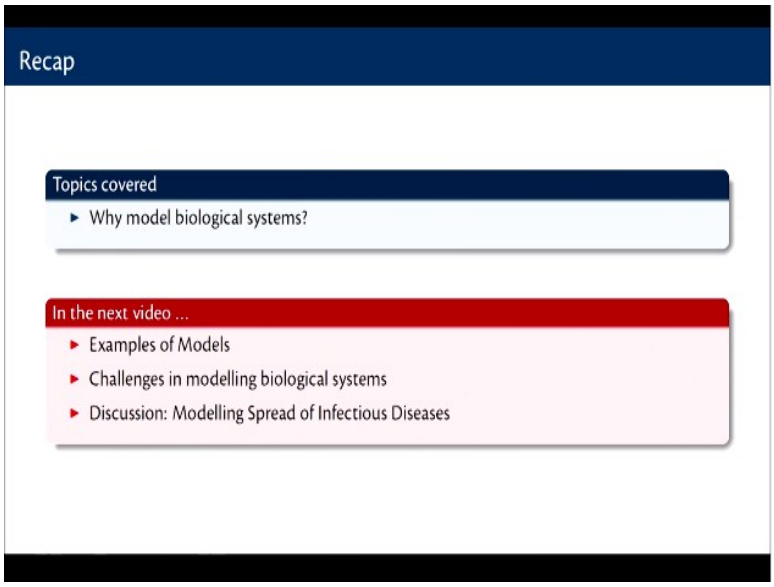
(Refer Slide Time: 22:29)



And of course, if you fail to see the big picture. Every country has their own fable about this blind man and the elephants. So, if you do not look at the big picture, if you don't look at the whole system, so this has been argued for systems thinking for a long time and it is obviously therefore extends to systems biology wherein if you do not look at the whole elephant you might think it is a rope or a log or a snake or whatever.

So depending upon which part of the elephant you look at, you will have a completely different observation and this holds for every cell that you will be studying as well. If you only look at a particular part of the metabolism, you might come up with the completely different conclusion than what you will be able to arrive at when you look at the whole system. There are these kinds of challenges when you model biological systems.

(Refer Slide Time: 23:21)



So, in this vide we have seen what are the motivation for modelling biological systems. In the next video, we will start with some simple example models, look at some of the major challenges in modelling biological systems and have some discussion about modelling the spread of infectious diseases which is a motivating example for modelling a biological system.