GEORGE CHURCH: OK. Well, even though we've been hinting at networks pretty heavily all the way through the course, these are the three lectures where we actually take it on. We really started this at the end of last week's lecture, so-called protein 2. Where in the process of talking about protein modifications and quantitating metabolites and their interactions with proteins, we started talking about the sorts of sources of data that you would have that would allow you to get at a quantitative analysis of protein networks, such as red blood cell. So we're going to pick up on that theme by talking about macroscopic continuous concentration gradients, and then contrast that with mesoscopic or discrete molecular numbers.

We're just going to very briefly touch upon the issues in discriminating between the stochastic modeling and the continuous modeling. And it's a very interesting connection between the red blood cell model, where you have a few examples of cooperativity with modest Hill's coefficients-- to take that to an extreme case where you have actually bistability, where you have two stable modes which are separated by a very cooperative interaction. And then we'll talk about copy number controls, another opportunity for doing modeling either by macroscopic continuous modeling or this stochastic modeling. And then, after the break, we'll talk about flux balance optimization, which I think is a really exciting and clever way of leveraging the little bits of information that you have about very complicated regulatory networks, biochemical networks.

So I think we've kind of just barely touched upon this before. The particular networks that are being studied, how did they come to be studied? Why and how?

And typically, what they have in common is that they have large genetic and biochemical kinetic data sets to go with it-- and/or. Right now there is no model that describes all of the interesting aspects of an entire cell or an entire organism. Usually, there are little pieces of it. The closest to a whole cell is the red blood cell. And this is because there's no biopolymer components to it, no biopolymer synthesis.

Today, we'll talk a little bit more about these two related topics, which is cell division cycle and the segregation of chromosomes during the cell division. Here, the key point is the critical nature of single molecules in these that come into play in the dividing cells. And in this, they'll be talking about bistability, how there's a decision of either to take the next step in dividing the cell or not. And how that bistability can be achieved either stochastically, where you're dealing with the fluctuations of single molecules greatly affecting a switch, such as a phase lambda switch, or you can have it involving a large number of molecules, where stochastic seems to play less of a role.

And then, at the end, we'll talk about how we can do comparative metabolism, where we literally integrate genomics with a network model of biochemistry at many different levels. Of course, there's the genome encoding the components of that network. But there's also the systematic knockouts of genes and their effects on the network.

Now, this slide is also review. But it puts in context where we'll mainly begin talking today. We'll talk a little bit about ordinary differential equations, both for-- at the beginning, under red blood cell, and in the bistability discussion, how you can get, even with just concentration and time, you can get these very interesting behaviors, highly cooperative. And then we'll drop concentration and time by a steady-state approximation when we talk about flux balance in the second half.

And of course, before, we were talking about molecular mechanics in the context of protein structure. And master's equations are the way of looking at stochastic single molecules, which we'll mention in passing. Eventually, in the network, discussions will get to spatially inhomogeneous models, where you actually care or realize the importance of where particular molecules lie in terms of their function.

Now, what are the limits and problems in connecting the in vitro parameters that were so key last time and this time in developing a system model? By isolating particular molecules, you can cut off, make the system simpler. But then there's the problem of reintegration.

And here, this is more historical artifact than really critical to this discussion. And originally, enzyme kinetics would ignore the products for technical mathematical reasons. But as you can see from last time, we showed that you could represent those equations quite well and even make the measurements. In the presence of products, it would just be another few terms in the rate equations. More critical, however, is that including the product along with substrate in the measurements and the modeling is just the first step, because how many different products and substrates and other regulatory molecules might be involved that you don't know about initially?

In addition, the conditions for doing the in vitro measurements, it's hard to do them at the concentrations in cells. And in cells, they're nearly crystalline densities, as I say, up to 30% or so, which is basically the very high concentration of solute that occurs-- proteins and other macromolecules. And the substrates, the small molecules, are typically in vast excess in in vitro reactions. But they're very close to equal molar in vivo because a lot of them are bound up with enzymes.

And you get interesting observations, such as this one mentioned at the bottom where a chemical reaction, which is spontaneous in solution, which is the epimerization of galactose, does not occur in normal cells of E. coli unless they have an enzyme that catalyzes this normally spontaneous reaction. So this is curious that it happens in solution. Doesn't happen in the cell unless you have the enzyme.

OK. So with all those caveats in mind, and recognizing that even though this model has more measured parameters than almost any other cellular model, they were all measured in vitro, with the caveats that we just mentioned. And the cell is aimed-- the function of all these network is mainly to provide the redox that keeps the hemoglobin reduced and the ATP that keeps the osmotic pressure under control. And also, the cell is-- even though we've shown a little structure in this network, the structure is assumed to be fairly continuous within the cell, which is a fairly good approximation in this case.

So this is not merely the stapling together of some kinetic in vitro parameters. We have other considerations in a real cell. We have the sort of physical parameters of the mass balance, energy balance, and redox balance. We've already mentioned energy and redox. But you have to have conservation of mass, as we'll see develop quite a bit more in the second half of this talk.

In addition, there are physics, such as the osmotic pressure and electron neutrality. There are cells which do have transience of non-electrode neutrality. But for the red blood cells, certainly, one of the goals is to maintain very close to electron neutrality and osmotically stable. You want to have as many non-adjustable constraints as possible.

As in other modeling systems, if these are measurements, rather than adjustable model parameters, then it allows you to test the few hypotheses you have, and have them overdetermined, and look for contradictions, outliers. And then, eventually, we'll see advantages to knowing what the maximum fluxes-- the maximum rates that you can have in these complex networks. And we could incorporate gene regulation, as we've seen lots of wholesale, increasingly accurate data on gene regulation is becoming available. It would be nice to integrate these, because the expression of the proteins affects their activity. The activity affects these fluxes.

So these fluxes are represented here in slide number 8 as dx dt, where each of the x sub i's represents one of these blue dots, a node in the network, where you have four basic processes that affect it-- up to four. You can have synthesis steps, degradation steps. So synthesis produces transport, can bring it into the cell. Degradation removes it. And it can be utilized, incorporated into the body of the cell.

And you can think of this as a sink removing it from the free population of-- for each x molecule. And this can be restated as a stoichiometric matrix [? so ?] ij, where that's mainly 1's and minus 1's. As you can see, in front of each of these fluxes, synthesis degradation and so forth will be 1's and minus 1's that refer to the stoichiometry, and sometimes 2's and 0's. And then the transport is its own vector.

And you'll see the utility of the stoichiometric matrix where i is the metabolite number, and j is the reaction number or the enzyme number for all the possible reactions that can occur in a cell. These can be all possible reactions. Then you can toggle them on and off with mutations or changing different cell types.

Now, this particularly rich system, the red blood cell, has been modeled many times and continues to be modeled since the mid '70s and now into the 2000s, starting originally just with glycolysis, later adding pentose phosphate, nucleotide metabolism, various [? pumps, ?] [? osmotic ?] consideration. Hemoglobin ligands have been treated from time to time, and less so issues having to do with [INAUDIBLE] and shape. No model includes all of these in one model, although it's really very close.

But there are models that include all of the metabolism that we know and the transport osmotic properties. Relatively few of those were made available broadly. But now they're increasingly being made available freely on the web, as models should be.

The assumptions behind this is that-- like I said, not everything is modeled. Some of it, for mathematical convenience, you will typically, in order to do differential equation-- since there's vast ranges of time constants, from things that happen extremely quickly to things that happen extremely slowly, typically, what you do is you'll model with a window in the middle, where you'll say, things that happened very quickly can be treated as a pseudo-equilibrium, as we've listed in this middle line here on slide 10. Things that happen extremely slowly can be treated as a constant. If they happen over the period of years, then in the course of an experiment that might be hours, they can be treated as a constant or something that you systematically explore.

Although we'll typically be ignoring little pieces of the metabolism, like [INAUDIBLE] metabolism calcium-- either as data comes in from other systems and try to treat it by homology or analogy. In addition, when we talk about a typical cell, this can mean that we assume homogeneous distribution of molecules within the cell and homogeneity from cell to cell within a particular organism. It also means there's a tendency to model a wild type without respect to polymorphism. Although in human population, with red blood cell function in particular, there's quite a literature on mutations that affect the functioning of the proteins within the red blood cells. Probably one of the best studied human genetic systems.

OK. Surface area is not absolutely constant, although for the time being, it's modeled as such. This is some examples of a subset of it. This is a subset that refers to glycolysis.

You can see they all have the same form where you have a change in a concentration in moles per liter with respect to time of some small molecule-- here glucose-6-phosphate. It's some synthesis rate with the subscript being hexokinase. This is the upper left-hand corner. And then that's the synthesis minus the sinks, the degradation rates, which go through two other enzymes.

Again, a reminder-- at the bottom, you'll see this come up a few times, just a reminder. Each of these is a form of change in the concentration of some small molecule with respect to time-- is the sum of the synthesis, subtracting the degradation transport utilization. OK.

Now, remember we focused in on one little piece of this a couple of times now. This is a step that happens to be allosteric. That means that, depending on the concentration of the-- the top part of this formula tends to be composed of substrates and products having an effect. And then this term in the denominator has a fourth power dependence on a variety of other site effectors, including AMP.

And you can see that the velocity is either hyperbolic, which is the upper curve-- which kind goes smoothly up and plateaus. While a sigmoid curve implies greater cooperatively, which can be affected by some of these other site effectors, which is more sigmoid in shape. And we'll use this as a stepping stone to talking about what are the various ways of getting that sigmoid shape?

We've already talked about how proteins can be multimers, dimers, tetramers, and so forth. That could be one way of achieving the sigmoid shape via a conformation change which senses the second site. But we'll talk about another way in just a couple of slides.

When we have the kinetic expressions here, they have the form of the previous one. Most of them are simpler than the one in slide 12. The model has a total of 44 rate expressions. They have about five constants, on average, so about 200 parameters. These are not truly adjustable in the sense that they're determined from the in vitro reactions.

What kind of assumptions? We've already mentioned the difference between in vitro and in vivo. We have the lingering question of how many effectors might there be that we don't know about?

Typically, these in vitro experiments were done with a small number of substrates and products that you know about. But as a worst case scenario, Mike Savageau you know, likes to trot out the glutamine synthetase, which fortunately is not in this particular model. But it could be that there's an enzyme that's just as complicated but hasn't been studied as much.

In the case of glutamine synthetase, there are three substrates. Remember, the previous example, there were only two substrates and two products. Glutamine synthetase, three substrates, three products, and has nine allosteric effectors, rather than the three or so in the previous example. So this gives a grand total of 15 different molecules you need to track.

So the number of different measurements you might have to make, hypothetically-- no one's actually made this number of measurements. But Mike Savageau likes to point out that even if you only did four concentration points in this multi-dimensional space, you have 4 to the 15th measurements, or a billion measurements. A billion isn't nearly as intimidating as it was when he made this statement in '76, but it still is not something that's routinely done.

What other constraints? There are these physical chemical constraints of osmotic pressure and electron neutrality, here stated a little more explicitly. You have pi i equal to pi e. That means the pressure on the inside is equal to the pressure on the outside. That sounds like a good way to balance things out so that the cell doesn't explode.

And explicitly, what that means is these gas constant r times the absolute temperature, degrees Kelvin, times the sum of the pressure components for the j molecule, going up to m chemical species, for i, standing for Interior, is equal to the same sum, equivalent sum for the subscript e, for Exterior. Electroneutrality has the same set of concentrations for the i, interior, and j molecules, where now z is the charge, where charge is the same z that we had in m over z for the mass spectrometry.

So, OK. Now, some of the models I give you today, we will compare it to the calculated and the observed, as we have done before. And here it's shown a little bit differently than how I've done it before and how we'll do it later on.

Typically, we would have observed on one axis and calculated on the other. In general, we're looking for outliers. And here we're sorting by the degree to which they deviate from observation.

And so the deviation is going to be observation minus the calculated. And the degree of deviation can either be normalized to the standard deviation-- which is basically normalizing it to how confident we are in the experimental measure. Or it can be normalized to the averaged value, which then becomes less dependent upon the accuracy of the experiment and more dependent upon-- with a fraction of the-- and so we've sorted it on the ladder. And you can see that most of them are less than two standard-- sorry. Deviation is less than twofold the average value and less than seven standard deviations in terms of the measurements.

But the ones that are furthest to the right are clearly the ones that require the most attention, either in the experimental measurements or in the modeling. These are steady-state measures. This is kind of an abuse of a beautiful kinetic model. But it reflects the limited data that exists. And it's much easier to collect steady-state data where, basically, the red blood cell-- every particular molecule, even though there are fluxes in and out of every molecule, the molecule concentration itself is staying constant.

So that's how you-- if you're assuming each molecular concentration is staying constant with respect to time, you're just measuring steady-state levels. But if you're more interested in looking at the dynamics, the movie of how molecules will change if you perturb the system, you can think of a wide variety of different curves that the timecourses can take. And then the challenge is how do you represent them?

You have 40-some different small molecules. You're tracking the concentrations. And then the timecourse can range over hours.

But one way of doing it is doing pairs of substrates at a time, substrate a and b, and then monitoring the timecourse as a vector. Think of these as a series of little points along here. And let's look in slide 17, look in the upper left, number 1.

If, let's say, a is converted to b, so that a plus b is equal to some constant, you can see this slope of negative 1. This is exactly negative 1, because for every molecule of a that's consumed, a molecule of b is produced. And so when you see this perfect slope of negative 1, then that's the kind of relationship you expect between those two, even if you're randomly sampling a dynamic system. Here you're taking a timecourse.

Number two is a pair of concentrations [? in the ?] equilibrium. You'll get an equilibrium constant, and the ratios here will not be negative 1. They will be some constant which determines that equilibrium.

You have two dynamically independent metabolites, as in quadrant III here. Basically, as you march through increasing b, a can stay constant because it doesn't really care. It doesn't respond to changes in b or changes that result in b changing. And then maybe some other set of dynamics will cause a to change, and b stays constant. And if you sample enough timecourse, you may find that this fills up the entire space concentrations available to a and b showing no correlation.

Another interesting type of phenomena you can see is that not every possible concentration of a-- and it's not completely independent of b. As you might start at a particular point in a time series at this gap here in the lower right-- and you start decreasing b and increasing a. And then, at some point, the dynamics of the entire network, not just a and b, contribute to now a taking a dive down and b increasing. And eventually, you return to that steady-state point, and you've described all the conditions that you might be able to achieve in this closed loop.

So we're going to look at these kind of phase diagrams. The concentration of a versus concentration of b-- where a series of time points that would be color coded. These can be either lumped, as they are in this diagram, or in the next slide, we'll see them separated out one metabolite at a time. But it's the same concept, whether you've got a group of metabolites involved in glycolysis lumped together, compared to, say, a group that are redox.

You have in the lower left of each of these quadrants is this time series that we've been looking at. And the upper part is the correlation coefficients, color coded so that blue is a negative correlation, and gray is very significant positive correlation, and everything else is something in between. So what you see from these is you see, for example, here's this up curve that's very close to the negative 1 curve, as if there's a conservation reaction here between glycolysis and the adjacent steps. You see little loops, for example, in the lower left, adenine biosynthesis, in that row, and so on.

You see examples of each of these kind of behaviors here, where you're going from the red point-- these little dots in red-- to green to blue to yellow to the end gray in increasing time coarseness, starting with 0.1-hour resolution and ending with 300-hour resolution. This is lumping, where we're kind of looking at things like ATP and redox loads. Now, if we look at it one molecule at a time, you obviously get a more complex-- you get every possible pairwise combination of molecules. And you can see these full dynamics.

Now, these are not data. These are all simulations. And unfortunately, we don't have that kind of dynamic resolution in experimental data just yet.

But this gives you some idea. If you see some particularly interesting phenomenon here, then it might be a motivation for going in and looking at the data in more detail. OK.

Now, we mentioned this difference between the ordinary hyperbolic curve-- in the lower right of this whole slide, the upper left of the insert-- and the sigmoid curve that you get from an allosteric interaction. And within-- this whole cell is set up for transporting oxygen. And as oxygen concentration increases for hemoglobin, you'll either get this hyperbolic or sigmoid curve as it gets increasingly sigmoid, increasingly cooperative, with increasing amounts of one of the intermediate metabolites, 2, 3 diphosphoglyserate. And you can see, this is now the connection between the glycolytic pathway. The regulation is sensing the state of the glycolytic pathway and the hemoglobin.

In addition, there are connections with the pH, which is also regulated in this, and the redox. You can see just above it here, the hemoglobin going to the unproductive methemoglobin state. So you can see there are connections in the network between this ultimate function, which is transporting oxygen, and all these intermediate metabolism components. And it also brings us to this topic, again, of this cooperativity and how does it arise here? And the hemoglobin has a tetrameric conformation state change, which is there's a second site that binds this organic diphosphoglyserate.

But another direction we can take-- so in the lower left-hand corner of slide 21 is the same icon again of how we get increasingly cooperative from hyperbolic to sigmoidal, to the point where this becomes almost vertical and displaced from the origin so that the cell, at a good point in response to a stimulus, will make a decision to commit to the next phase in cell division. This is similar to the cell division that we talked about earlier in the context of the microarray analysis of yeast cell division. Here we're talking about a Xenopus amphibian oocyte, which has nice large cells to do this kind of study.

Where you need to decide to come out of G1 and commit to synthesizing DNA-- in the S phase, in the lower portion of the circle, where you get, now, two DNA molecules. And once you're convinced that you've finished replicating all your DNA, only at that point, then can you commit to mitosis, another major decision. And then you get two cells, and you go back, and you complete the cell cycle.

The little timecourse at the bottom here should be reminiscent for you of the timecourse we had of RNA synthesis of various clusters of RNA. Here it's the DNA synthesis. You can see it ramps up in the red S phase and then ramps down in the metaphase due to the creation of two new cells.

But we want to talk about how do we make this as responsive as, say, progesterone or something that is signaling cells that might be waiting for long periods of time to complete the next step in a division cycle for an external hormone stimulus that it's time to start the next step in cell cycle? We want this to be displaced from the origin. You don't want it to be just flipping on and off irrespective of a stimulus. But when it does flip, you want it to go very quickly.

So how would we model this? So look at the upper right-hand diagram of slide 22. And you have a set of these oocytes kind of diagrammatically indicating their state. Their state determined by to what extent are they ready to commit to the next division state?

Here we can think of it as the biomarker is the state of phosphorylation of a protein MAP kinase. If it's phosphorylated, then it's committed to this division. And we can think of this as the black side of this gradient, going from white to black.

But if you grind up a whole population of these oocytes and measure the total MAP kinase phosphorylation as a function of increasing stimulus, S-- in this case progesterone-- the response-- that is to say, the phosphorylated state and the commitment to mitosis-- will gradually increase, as indicated in the gradient model. But that's if you ground up all the cells. But the other way of obtaining that result would be if each cell is making an all or none decision.

And what happens is the probability of a cell being in that all or none state changes with increasing stimulus to progesterone. And that is the lower model and is, in fact, closer to reality, as indicated by the experiment at the bottom part of the slide here, where you have-- this is a part of a concentration curve where you're increasing the stimulus progesterone. But at this particular stimulus, you can sample individual cells. There's enough of the cell that you can actually do proteomics on individual cells.

And the proteomics here is a western blot. We mentioned this a couple of lectures back. And here you can see the two states of the MAP kinase.

The phosphorylated state is the slower. Electrophoretically, it's the upper band in this diagram. And the lower band is the unphosphorylated state.

And you can see there's no example here of a cell which is in an intermediate state, where it has half and half or 40/60 of the two different protein forms. However, if you, as a thought experiment, took all those cells and mushed them up, and took this and ran it all in one lane, you would see a mixture. You would see all the intermediate states, and that would be a function of progesterone. So this is a warning, similar to ones that I've said before, that when you grind everything up and mix populations of cells or molecules, you need to be careful because different cells may be in different states, different molecules may be in different states. And the average behavior is not the same as the individual.

But now, that's only part of the lesson here. The cells are going through this all or none process. We can monitor it by single-cell proteomics.

But how do we model it? Well, in very abstract terms, the response here can be modeled as in this Hill coefficient, where you have a stimulus, s, some kinetic constant, k-- kind of like a Michaelis constant-- where it's basically, as gets s closer to k, it has a larger response here, effect on the response. And that's nonlinear because you have the exponent h. And the larger h is, the more nonlinear it is.

So let's say that h is 1 in this little schematic to the right here. It's hyperbolic. No sigmoid character at all.

In the case of hemoglobin and phosphofructokinase that we talked about in the red blood cell, it's more sigmoidal, like an h of 2.8, almost cube law there. And in fact, even within this system, one of the steps that we'll talk about in the next slide-- where you have a stimulus of a Mos protein and a response of this same MAP kinase phosphorylation response, it has a modest sigmoid Hill coefficient of 3, just like hemoglobin. But the overall response of MAP kinase to progesterone has an enormous exponent of 42. That means it's almost vertical, and it's displaced from the origin. How do we get that?

OK. So-- whoops. So here is a proposed model.

And it's an interesting snapshot in the inevitable evolution of a model from something very primitive, which might have just been the Mos effects on MAP kinase, kind of as a direct effect here, which as we said, has a Hill coefficient of about threefold-- sorry. Hill coefficient of 3, no threefold, because that's an exponential.

But overall, combination of two other factors that you have a chain of modifiers, each close to saturation, meaning that each one has a slight sigmoid behavior. And you've got-- sorry. It could be a hyperbolic function, such as, let's just say, this dotted line that's going up smoothly from 0, where it says "neither."

That would be the effect of, if you just have a normal enzymatic reaction with no allostery, no feedback, no ultrasensitivity. If each of those has a component and you're close to saturation-- meaning your substrate is very high. And so you're going as fast as the enzyme will go.

You have a chain of those. You can show through kinetic modeling that that will create a high sensitivity to the reaction. And that's what the furthest dotted curve close to the axis is, ultrasensitivity alone.

And then, here, you have progesterone going in, in the upper left, affecting a complicated rate. Now, these rate constants don't necessarily mean a unitary, simple kinetic step. They can represent something as complicated as going from Amino Acids, AA, to a particular protein, Mos. And then the reverse reaction, k sub minus 1, is the degradation of Mos back to amino acids, which is not a reverse of the same enzyme set by any means.

The next step, k2, is simpler. It's just Mos being phosphorylated by, actually, our friend MAP kinase in its [? phosphorous ?] state, producing the-- for catalyzing the Mos phosphorylation, which catalyzes another phosphorylation of another protein, which then positively stimulates MAP kinase itself. So this whole thing is a set of positive reactions. Each of the phosphoproteins increases the enzymatic tendency for each of the other phosphoproteins to be produced.

So you can see that this is kind of on a hair trigger. If any one of these phosphoproteins gets produced, then it'll increase all the other ones, and it'll be a very cooperative procedure. And that's what this, furthest towards the axis in the lower left-hand corner-- where you have this positive feedback alone causes this very great tendency to just jump up from 0 to very high response with very little stimulus.

Well, that's dangerous. You want it to be nearly vertical, but you don't want to be nearly vertical at 0 stimulus, because that's unstable. So you want to move it over. And that's where the ultrasensitivity comes in, when you have this chain of modifiers putting both together as a solid black line, where it's shifted over so you have-- this assay was done with Mos, rather than progesterone, as the input. But you can see the overall increased cooperativity and shifting to the right.

So that's an example of how you can get this very high Hill coefficient and where you can get bistability without stochastic. You can imagine that you can have stochastic bistability. If you have one molecule in the cell, and either it's there or it's not, then you have bistability.

You have two states for the cell. Either it's the cell with the molecule or without. But here you can see that even with a very large cell-- Xenopus oocytes being one of the largest cells-- and very large amounts of proteins-- enough that you can easily see them in proteomics-- you can still achieve bistability with the right kinetic model. Not every random model would achieve that high Hill coefficient.

OK. So we're just going to briefly mention the other way of getting bistability, which is via stochastic, so small molecules. And here, an example-- so instead of dealing with very large cells with very large numbers of molecules involved, here in, say, bacteria in particular, a phage-infected bacteria, you generally have the case of very small bursts of activity.

A transcription factor will bind to a promoter. Before the transcription factor comes off, it may cause a small burst of a couple of RNA transcripts being made by a couple of RNA polymerases seeing those transcription factors. Then each of those RNAs caused its own bursts of protein synthesis where a whole series of ribosomes will bind in a polysome. And you'll get this double burst of RNA and protein.

And the stochastic binding of that transcription factor that starts that burst can be modeled by reasonably measured parameters for each of these steps. And you can see that cells one, two, and three, in the lower left of slide 25-- where time is the horizontal axis up to, say, 45 minutes or a cell division or two. While the number of product proteins here, measured in dimers of protein, fluctuates, where cell one gets an early start, early burst, and cell three hasn't quite hit its burst yet.

So you can see there's a lot of variation. And this is one way of achieving a bistable switch. But as you've seen-- not the only way. You can also do it where all the proteins are present in all large amounts.

If you do choose to go the stochastic route-- and this might be an interesting project for some of you. It's by no means shown to be mission critical for the community of systems modelers. But many people believe it is a way to go.

There has been great progress since 1977 when Gillespie proposed the algorithm named after him for stochastic simulation, a couple of chemical reactions in general, not just biological, biochemical reactions. Since then, Gibson and Bruck, within the last couple of years, have come up with an algorithm, which is now time proportional to the logarithm of number of reactions, rather than the number of reactions. Any time you go from n to log n, this is big progress. And this is done by better tracking of calculations that you can reuse. And so I encourage you to take a look at this aspect of stochastics.

Another aspect is people often think of the stochastics as kind of a nuisance. They increase the computer time that it takes to do simulations. They increase your uncertainty about the simulations that you then produce.

But there is an aspect of it which is just beginning to be harnessed in various fields of engineering, and biological engineering is no exception. And I give you two examples here to just whet your appetite. Again, I'm not going to go through them.

But you can see that you can actually make switches and amplifiers for gene expression-- gene expression being one of our favorite topics in this course-- which are based on noise and where you can get bistability using these fluctuations. And that's not too unexpected based on what we've just been saying. But in addition, you can get stochastic focusing where the fluctuation allows enhanced sensitivity. OK. So I encourage you to look at that.

Now, a particular place where you might worry that stochastics is coming into play quite a bit is in chromosome copy number, whether this is eukaryotic chromosomes or in the case we'll illustrate, a very simple case of plasmid chromosomes. Now, the interesting thing about plasmids is they can either be in lockstep with cell division, the way that eukaryotic chromosomes are-- in the cases of Xenopus oocytes we just talked about, where it makes a big decision. And that's the case of the R1 plasmids.

Or it can be more of kind of a cloud of copy number, where they're trying to be close to a target number where you'll have more copies than one per cell. And so as the cell divides, it kind of randomly takes a partition of that number of plasmids. And ColE1 is an example of that.

And you model it in order to determine the factors that govern it. This has implication. The copy number will affect the expression levels.

And the expression levels are of importance to biotechnology. And plasmids are, of course, also important in pathogenesis. The copy number affects pathogenesis since plasmids are a major way that drug resistance elements are passed around.

So let's take a look at one, hopefully, highly oversimplified version of this. Here you have two RNAs, imaginatively called RNA 1 and RNA 2. We'll start-- RNA 2 is transcribed here, on the bottom strand, from right to left. So actually, look at the very bottom of slide 30.

The magenta RNA polymerase is making RNA 2. And if nothing binds to RNA 2, if RNA 1 does not bind to it, RNase H will cleave it. It will then bind to blue DNA polymerase and will start replicating the plasmid.

On the other hand, if RNA 1, which is made on the opposite strand of RNA 2-- it's this antisense story. It will then come and bind to RNA 2, sort of in trans. It acts as a transacting inhibitor.

It's aided and abetted by the Rom protein. And now you don't get cleavage of RNase H. And so DNA polymerase doesn't have a primer upon which to act, and you don't get replication.

And this is, of course, not just a yes or no thing. This is something that's regulated and allows it to feed back to get the right copy number. You don't want it to get an infinite copy number, or else it'll choke the cell that's harboring it. But you don't want it to drop down low enough so that then many cells will segregate with no copies of the plasmid.

So you want to have a mass balance. You have to have conservation of mass. You want to be able to model both the initiation and degradation and inhibition. So you do this by making some simplifications that the RNase H rate is fast. Remember we had slow and fast reactions that we would eliminate, so too with the DNA polymerization.

By subsuming the RNA 2 concentration in an RNA 1-based model, you can simplify it so that you're really only considering two species, RNA 1 and the plasmid DNA itself. We'll call these r and n for the two different molecules. So this is just a way of introducing a two-species model.

We're going to come up with a rate equation for change of RNA 1 with respect to time. And then the next slide will show the change in the concentration of the plasmid. Concentration of the RNA is r. Concentration of the plasmid is n. We have dr dt and dn dt.

And this is very simple. It's just like what we were talking about with the metabolites. You synthesize the RNA. That's a positive term.

You degrade it, or you dilute it out. The dilution is based on the growth rate. Mu is a typical-- it's used in growth rates in population genetics, and here it's used in chemical kinetics. In fact, this is, in a certain sense, an example of a very exciting field where you're bringing together population genetics and chemical kinetics into one place. And population genetics and chemical kinetics, when they come together, unite some of the most disparate parts of this course.

OK. So now we've got an equation here where the positive term is k1, the rate of initiation of RNA synthesis. And it's, of course, the more molecules, n, the higher the concentration of the plasmid, then the more RNA you're going to make. So that makes sense that you have the product, the rate constant times the number of plasmids.

Similarly, the loss of it is going to be related to the number of RNAs. More RNAs you have, the more that you're going to lose, more that you have to lose. So that's the RNA, and this is for the DNA.

Here you have dependence on the RNA. RNA 1, remember, is the thing that we modeled in the previous slide-- is an inhibitor. And so it's going to, when it binds to RNA 2-- which is implicitly modeled here-- it's going to have this inhibitory constant in the denominator. And so as the inhibitor RNA goes up, this inhibitory term goes up, the forward rate goes down. And the rate of replication is going to be also dependent on plasmid copy number, so it goes up within.

The dilution rate is, of course, dependent on n as well. So the idea of this, in the next slide, is going to be to solve for the plasmid number. So slide 34, we have how you would implement those two equations that we had in the last two slides-- are shown on the very top left part of the slide.

dr dt is abbreviated dr here in mathematical format. It's that same k1 constant times n is the concentration of plasmid molecules. Then the negative is the degradation rate. And the dilution by cell division, mu. And that's times r, which is the concentration of the inhibitory RNA 1.

And in an [? analysis ?] equation, which we've already seen before for the plasmid molecules, n, the change in concentration of n is a function of time, dn dt, here abbreviated dn, is there. And then we're going to solve it.

So these first three things are setting it up and asking the program to solve it. And we'll do it under the constraints of dr dt is equal to 0, dn dt is equal to 0. You will recognize this as the steady-state assumption. Even though there are fluxes in and out that are non-zero, the net effect is zero. And so that's the formula for steady state.

We're going to start at a dilution rate of 1, where you'll have some steady-state level. And then we're going to watch the dynamics as it goes to a dilution rate that's slower, that is to say, the growth rate is slower. And as the growth rate is slower, you'll expect, maybe, to accumulate more plasmid molecules.

Because they're not in lockstep with the cell division rate, the cell grows more slowly. And there's not some other feedback, and we haven't put any other feedback in this model, then it should go to about twice the level. So you do the symbolic solver in the top here. And you get this symbolic solution here.

And if you do the numeric solver, NDSolve, then you get a very similar solution. And you could plot it. Here you're plotting y, which is a plasmid copy number from over a time range of 0 to 3. And you can see it goes up from slightly over 1 to slightly over 2 in terms of a concentration of plasmid, as you might expect from lowering the dilution curve.

Just as we had stochastic models for the bistability that we talked about earlier-- the Xenopus bistability could be continuous. And the lambda model could be stochastic. Here there are stochastic models for Copy Number Control, CNC, which are very interesting. I urge you to look at them-- where you can have, basically, using stochastic modeling to do molecular clocks, where you can reduce the rate at plasmid loss. You can see, in that last one, if you had a very small number of molecules, you would have a loss in some of the cells that would be more accurately modeled in a stochastic model.

Now, we want to go from these models of red blood cell where you have metabolism without polymer synthesis, and the CNC model where you have polymer synthesis of RNA and DNA, but without metabolism, to a more integrated cell where you have both going on. And you want to represent the full optimization that must occur, getting metabolism optimally suited so that you get the right kinds of macromolecules made in a complicated cell like E. coli, which can adjust to a wide variety of different growth conditions. So what are the problems here?

The number of parameters that we needed and had for the red blood cell was enormous. It was 200. It was a tour de force to get those. For E. coli, it's orders of magnitude more that are needed. Because instead of 40 enzymes, we have somewhere between 400 and 4,000.

So measuring parameters is a problem. And we have the same problem about in vitro versus in vivo. We have the same set of constraints. And we want to focus more of our attention now on the flux constraints.

So after a short break, we're going to come back and talk about the flux balance as a solution to this. And just as we had with the red blood cell, we're going to be focusing in on ways to relook at the way we think about the synthesis and degradation of the molecules in this network to see if we can rephrase in a way where we can ask interesting questions about the optimization of these systems. So take a quick break. And the second half, we'll talk about the flux balance.