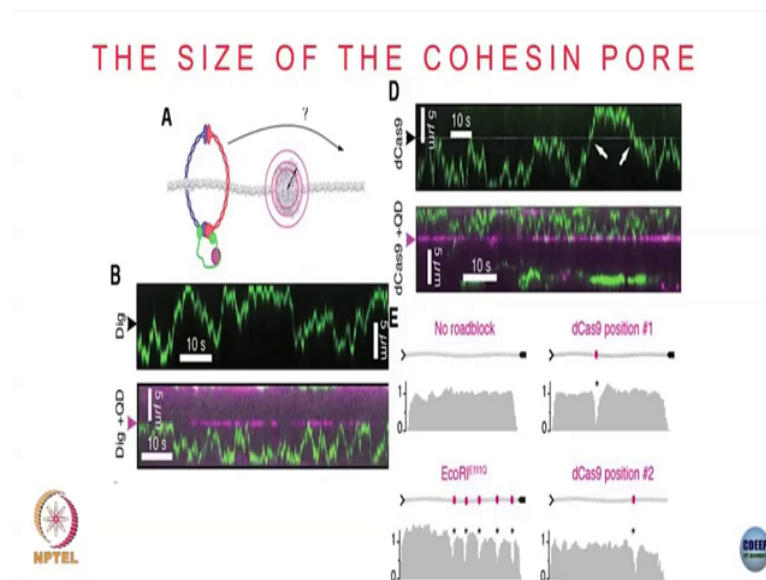**Lecture – 34**
Nucleosomes as Barriers, Hi-C, and Contact Probabilities

(Refer Slide Time: 00:16)
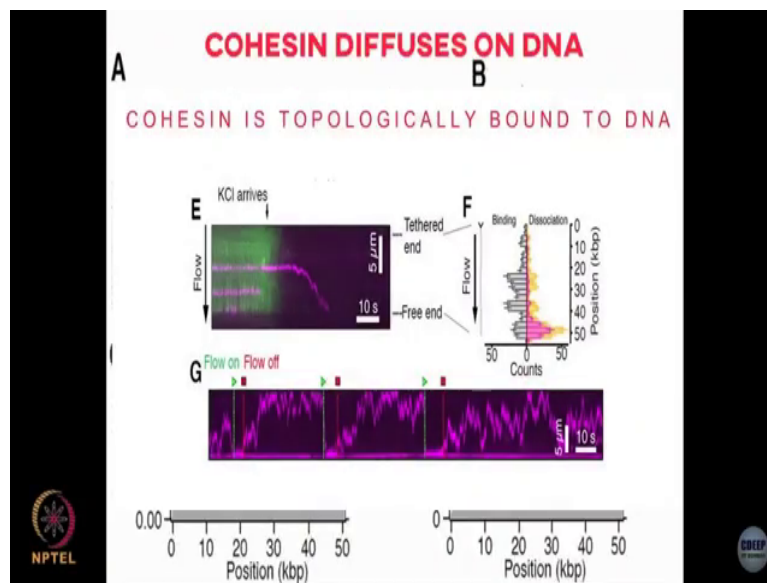
Anyway so, again they have done experiments, but not only have they put Nucleosomes you could put different Barriers now, barriers of different sizes.

(Refer Slide Time: 00:24)

So, what you do is that you put some sort of a barrier and then you ask how does it affect the motion of this cohesin along the DNA backbone, right. And now I can put barriers of different sizes and ask that well, for very small barriers you would expect nothing happens; for very large barriers you would say that it would never cross the barrier and do we see that or not.

So, for example, here is a very small protein, which I have put at this location on the DNA. So, if my whole DNA is around 50 kilo bases; at the 20 kilo base mark I have attached a protein and then I see what the cohesin trajectory show. And it because it this is very small, so this protein is around if I remember correctly around 2 nanometers; because it is a very small protein, it does not this cohesin ring does not even see this protein, it continues to do its random walk, happily on it is own, right.

But now what happens, so now what is done is that on the same protein you put a quantum dot, which is sizes written somewhere. So, it is a much bigger obstacle now. And then you ask the same you put it at this location, the location of the arrow; so everywhere on this there is this obstacles or rather at all times there is this obstacle at this location. And then you ask what happens to the trajectory and as you can see it goes, it hits this obstacle and then it comes back; because its large it cannot really cross over, it comes back it does random walk again, again when it reaches the boundary it has to turn back, ok.

By doing this in a control and you can do the same thing with different obstacles; by doing this in a controlled fashion you can say that, what is the effective size of this ring, right. When an obstacle becomes much big, so big that this ring cannot cross over this obstacle; I say that is gives me an upper limit to the size of this ring itself, right. So, they do this and define that this cohesin pore is roughly between I think 10 to 20 nanometers.
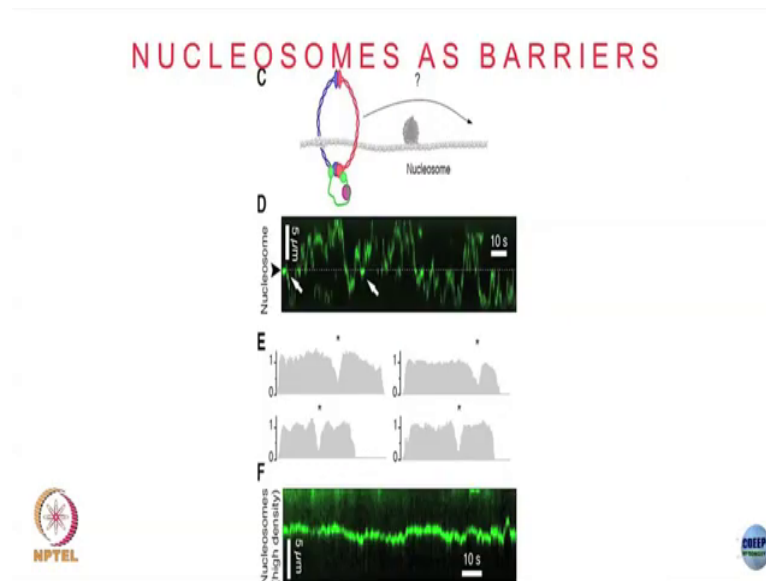
What happens if you actually put nucleosomes? So, this is a nucleosome. So, remember this is this DNA wrapped around this histone complex. And what you see is that; so here is where I have place this; this is a single nucleosome that I have placed at this location on the DNA backbone, right and I observe the trajectories.

So, what I will see is that, there are crossing. So, occasionally when it comes here, it crosses over to the other side; but also sometimes when it comes to this boundary, it does not really cross, it comes back over here ok; more often than you would expect purely on the basis of a random walk.

So, what this is roughly like a semi permeable barrier. So, I have this a one dimensional random walk of this cohesin protein on the DNA backbone and somewhere over here I have placed an obstacle, right.

(Refer Slide Time: 03:19)
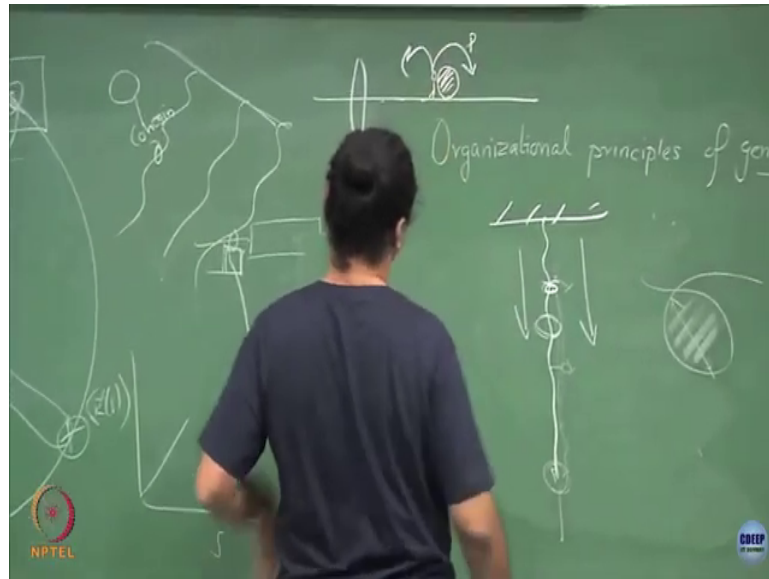
So, when this random walker comes and meets this obstacle, with some probability it will cross over; but with some probability to go back. So, it is this crossing over probability is not zero, but it is pretty small. If there was no obstacle here of course, both sides would be half and half. So, it is definitely not half; it is much much smaller than that, but it is not zero, ok.

And you can place this, so I do not know if you can see this figure; you can place this nucleosome at different locations and you can see that the permeability takes a dip exactly at

the location of the nucleosome. If you place the nucleosome here, the permeability takes a dip here; if you place it to some other location, it would take a dip there.

So, it functions it sort of obstacles, not complete obstacles in that does not completely stop the motion; but it definitely slows down the motion. And you could ask that well; therefore, what would happen, this is experiments for a single nucleosome. What would happen if I put more and more and more nucleosomes? So, they put many many nucleosomes now. So, this lower figure is for an extremely high density of nucleosome.

So, everywhere along here there are many many nucleosomes and then you look at the cohesion trajectory. And as expected because of this extremely high density, you see that the cohesin does not really move at all; it stays fixed at a roughly fixed at a particular position on the DNA, right, because whenever even if it moves a little bit, it sees another obstacle and so on.

(Refer Slide Time: 04:57)
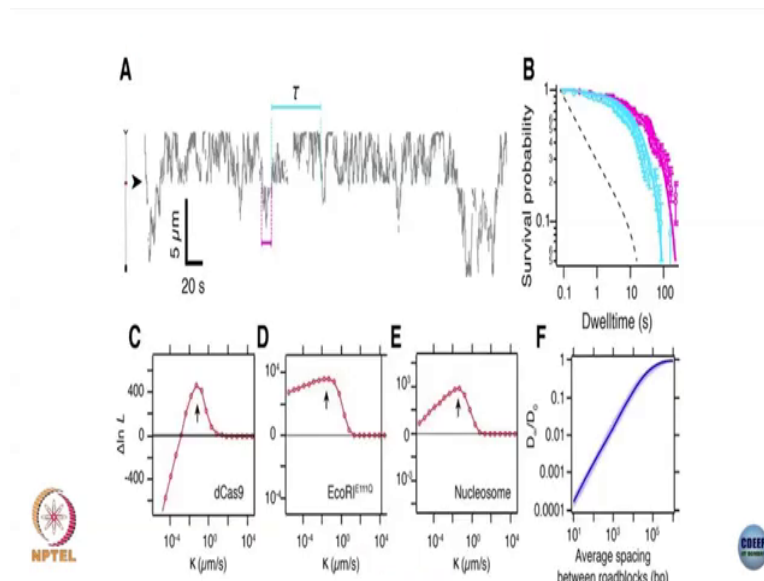
So, this is sort of, so you can estimate how the you can calculate various things like survival probabilities and so on; but you can estimate how the diffusion constant would sort of fall, as you put in more and more obstacles, so or as you decrease the linker length between two nucleosomes.

And the answer that you get, where the answer that these people get is a diffusion coefficient, so realistic link length like I said is roughly 30 to 100, 150 maybe, so somewhere along this region of the curve. And there you would get diffusion coefficients if I took that, for completely bare DNA, it is around 1 micron square per second; you get roughly of the order of 10 to the power of minus 3 minus 4 micron square per second.

And this is actually a puzzle; in that it is not clear that if you have an object which is diffusing on a 1 D lattice with something like 10 to the power of minus 4 micron square per second.

And then you ask that well I know in high c; I know from these experiments I will talk what experiments in a bit.

(Refer Slide Time: 06:01)
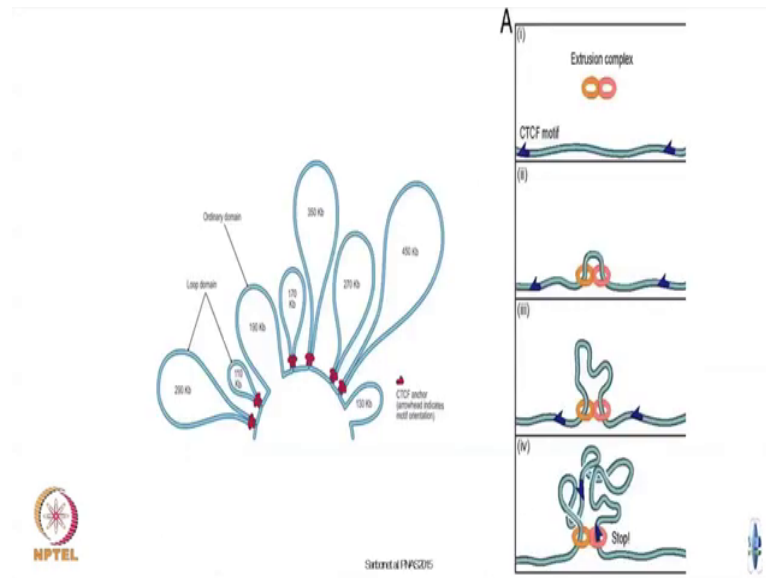
But I know from these experiments that I form these very large loops of the order of kilo bases and even mega bases. I could ask that given this diffusion coefficient what would be the typical time required to form this large loop, right. And it turns out that if you put in this diffusion coefficient number and if you put in this typical sizes of the loops; the timescales that you would get are of the order of hours.

Whereas typically we would need to form these within less than a second right; because the cells you would have to organize the genome very quickly. So, we need to form these loops on the order of seconds; but it seems if you put in these numbers which are measured, it seems

like it would take roughly a hours or something like that minutes or hours at least to form these large loops.

So, again this is something that is not completely understood; we understand how these loops form, because of these loop extrusion factors like cohesion. But we are not sure how to sort of reconcile the timescales that, we get from here and the timescales that we observe in experiments, ok.

(Refer Slide Time: 07:07)
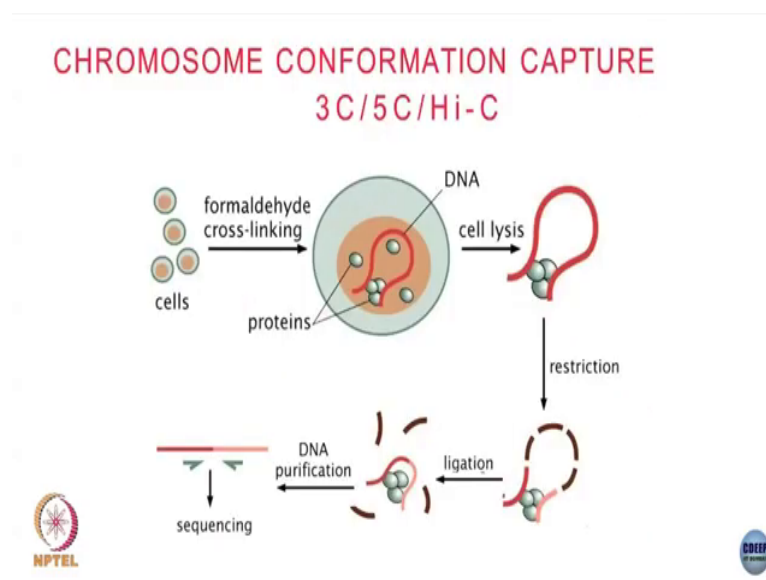
One more set of experiments; one more sort of set of experiments. So, these are a very nice class of experiments called chromosomal confirmation captured and again these tell us something about this large scale statistical properties of the chromosome, ok.

So, what I want to ask again is a related question that; so one thing that I asked was from these fish experiments one thing that we could measure was the average separation between two elements on the genome which are S base pairs apart. I want to ask something in a similar spirit which is that, if I take a chromosome or if I take any polymer; purely randomly for example, two segments of this polymer could come very close to each other, right. If these two segments were very close together along the backbone, they would have a high probability of being found together; if they were very far apart, they would have a low probability of being found together.

So, that is something that is call the contact probability and it is denoted by p c of S. So, this asks that what is the probability where two base pairs, which are S distance apart on the backbone; has come in close proximity to each other in 3 D space. So, that is call the contact probability. And there are these very nice experiments called chromosomal confirmation capture experiments, which measures this contact probability. So, I will just give a quick run through of what they do.

So, here is my DNA; so here is my, right; and let say that this region and this region are very close to each other. What I do is that I cross link these regions, so that they stay close to each other; then I lies the cell, so that I can take out the chromosome; I then chop off the chromosome into small bits, ok. So, this region which was sort of very close to one another, stays in one segment and these remaining things are chopped off. Then I ligate, so I wash away all these other strands which do not have this protein marker; and then I ligate these two strands this red and the yellow, remember which were initially very far apart. So, let say this was red and this was yellow.

So, they are initially very far apart along the chromosome; but at the moment I did this experiment, due to whatever reason they were close together, whether it was due to some loop exclusion factor or purely randomly or whatever and they were close together. And so, I freeze I sort of freeze them in that snapshot in time and then I do this like I do this restriction and this ligation, so that I join them together and then I wash away this proteins and I sequence this DNA that I have, right.

And this DNA now contains one segment which is from this region the red part, another segment which is from that region which was the yellow part. I do this across millions of cells at many many times and I get an ensemble of experiments which tells me that; how often do I find this segment in this segment, which was some S base pairs apart, how often do I find them in contact, right.

So, if I have done a million such experiments and in thousand of them I find that these are in contact like this; I would say that, that gives me my contact probability averaged over all S and so on and so forth. So, you can build these curves now and you can do this on actual genomes. So, for example, here is an x contact probability, your normalized contact probability as a function of this genomic separation along the backbone for the yeast genome and for the human gene. And here is what it looks like; it falls off roughly as a power law and so it p c of S goes as S to the power of minus beta.

And for example, in yeast roughly that b appears to be minus three halves on an average; in humans that beta appears to be sorry since I have written minus beta, beta appears to be three halves in yeast and one in humans. This sort of tells you a little bit also about how tightly the chromosome is packaged, right. If you did it for different chromosomes 1, 2, 3, 4 whatever up until 23; it would tell you if there is a difference between how tightly these different chromosomes are packet.

(Refer Slide Time: 11:49)
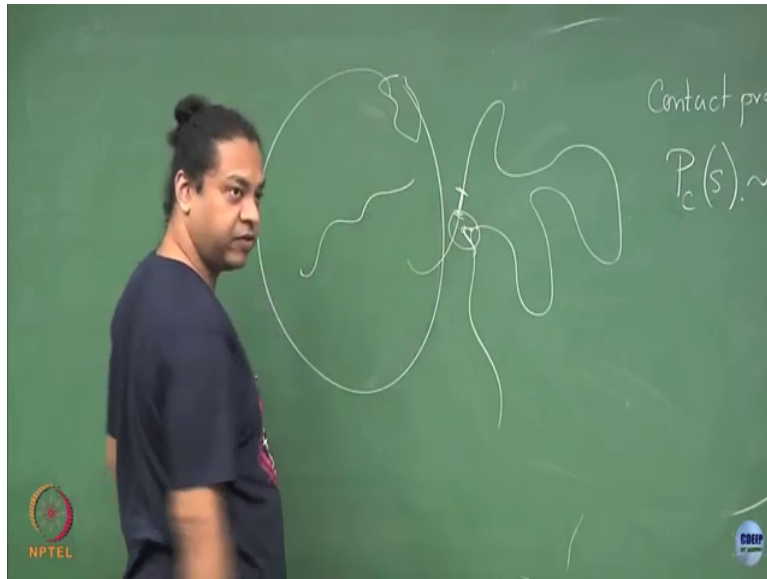
For example in that figure I showed earlier, the chromosomes that were on the periphery something like this versus something like this; you would see different contact probability curves for this chromosome versus this chromosome, right.
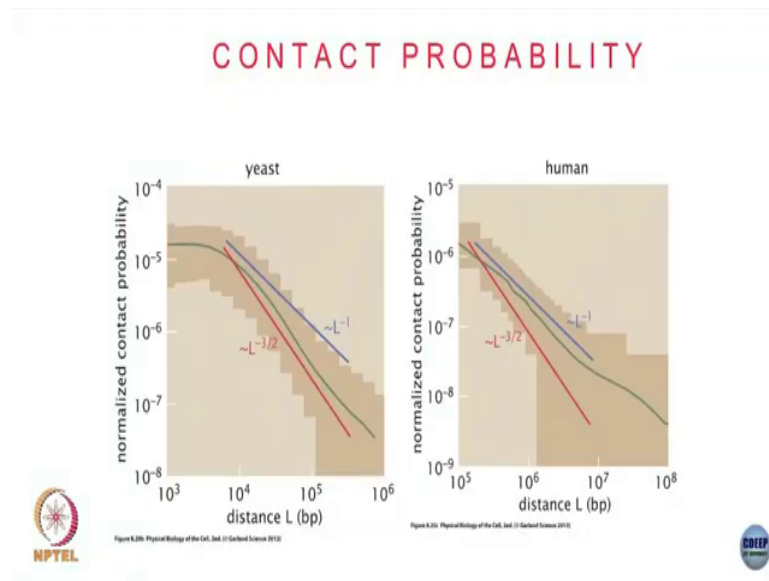
And this is something that you can actually calculate from simple polymer models. So, let us just see. So, this is in humans remember; humans gives S to the power of minus 1; yeast gives S to the power of minus 3 halves, ok. So, we can ask that well; now I know at least one simple model for a polymer which is this random walk polymer. If I were just looking at this sort of quantity; what is the contact probability as a function of the separation; what does the random walk model a random walk polymer tell me, ok.

(Refer Slide Time: 12:25)

CONTACT PROBABILITY - 1D RW

$$p_0 \; = \; \text{prob. of loop formation} \; = \; \frac{\text{No. of looped conformations}}{\text{Total no. of conformations}}$$

$$p_0 \; = \; \frac{1}{2^N} \cdot \frac{N!}{\left(\dfrac{N}{2}\right)! \left(\dfrac{N}{2}\right)!}$$
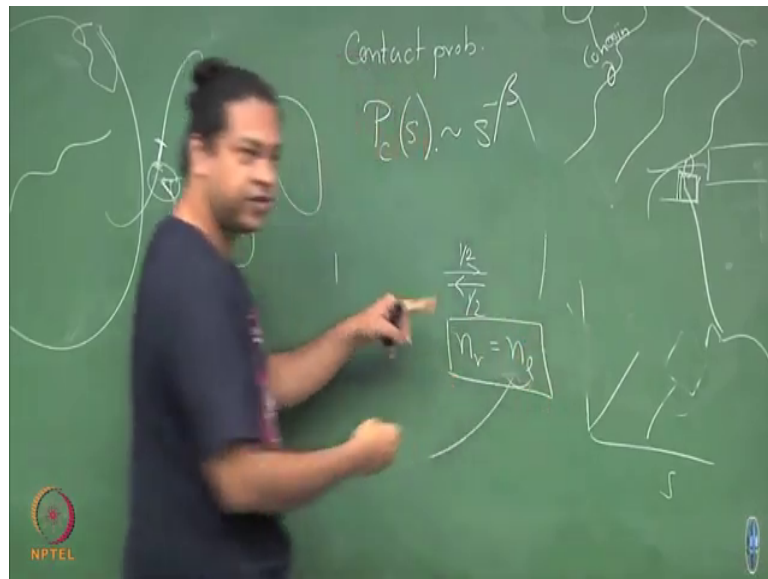
Use Stirling's approximation: $N! = \left(\dfrac{N}{e}\right)^N \sqrt{2\pi N}$

$$\Rightarrow p_0 \; = \; \sqrt{\frac{2}{\pi N}}$$

Cyclization probability $p_0 \; \sim \; N^{-1/2}$ for 1D RW

So, here is my probability of loop formation. It is the number of loop confirmations by the total number of confirmations. So, let me say if this was the 1 D polymer; let me just again do an idealized case first.

So, if this was a 1 D polymer, remember my 1 D random walk can take steps to the right or to the left with probability half and half, let us say if it is an unbiased random walk. And then a looped confirmation would correspond to, when the number of steps it has taken to the right is exactly equal to the number of steps it has taken to the left, right. So, it comes back to the origin.

So, I can calculate the probability of that. So, that is just N c N by 2 that is the number of loop confirmations and the total number of confirmations is 2 to the power of that. I can use because these ends are very large, I can use Stirling's approximation to simplify this and this will give me some answer. So, for example, if you use Stirling's approximation, what you get for the probability of loop formation for this one dimensional polymer, one dimensional random walk is it goes as N to the power of minus half.

Now, of course, I did this in the discrete sense; we could also write down the full probability distribution like we wrote down the probability distribution for the end to end vector is this Gaussian; e to the power of minus R square by 2 N a square. And let us say that I have some cut offs; that when two monomers approach each other within some distance delta, I will say that a loop has form. And then I can calculate what is the probability of a loop formation, so I take this probability distribution; I integrate over all N to N vectors which lie between minus delta to plus delta right and that will give me the probability of loop formation in this 1 D polymer.

And again you will see, if you work this if you put in this and you do this integrations from minus delta to delta, the simple Gaussian integration; what you will get is that, the contact probability will scale as N to the power of minus half, so S to the power of minus half; S being the genomic separation. So, for a 1 D random walk polymer you get p c of S goes S to the

power of minus half; but of course, these polymers real polymers are not 1 D, these are 3 D polymers. But that is fine, we know what is the probability distribution for a random walk polymer in 3 D, that also we did last class.

(Refer Slide Time: 15:01)

CONTACT PROBABILITY - 3D RW

The probability distribution for the end-to-end vector R in 3D is:

$$p_{3D}(R, N) = \left(\frac{3}{2\pi N a^2}\right)^{3/2} e^{-3R^2/2Na^2}$$

$$p_c^{3D} = \int_0^{\delta} 4\pi R^2 p_{3D}(R, N) dR \approx \left(\frac{3}{2\pi N a^2}\right)^{3/2} \frac{4\pi \delta^3}{3} \approx \sqrt{\frac{6}{\pi N^3}}\left(\frac{\delta}{a}\right)^3$$

Contact probability $p_c^{3D} \sim N^{-3/2}$ for 3D RW

A simple random walk calculation explains Yeast data!
What sort of polymer model would reproduce human data

So, let say I could take a 3 dimensional polymer now; this is my 3 dimensional probability distribution e to the power of minus 3 R square by 2 N a square with this normalization factor. And again I say that I 2 monomers I will say that a loop has formed; if 2 monomers approach within some distance delta of one another and therefore, I can calculate what is the probability of loop formation or this contact probability in 3 D. So, this 4 pi R square times this P 3 D D R from 0 to delta, right this if this delta is my cut off.

Again you can do this integration there is an R square e to the power of minus R square sort of an Gaussian integration. If you do this integration what you will; all I am interested in, I am

not interested in all this delta and A and so on; I simply want to see how it scales with the separation right, along the base pairs. So, I want to see how it scales with N. So, the contact probability for a 3 dimensional random walk scales as N to the power of minus 3 halves. So, if I put in this probability distribution for the 3 D random walk.

(Refer Slide Time: 16:15)
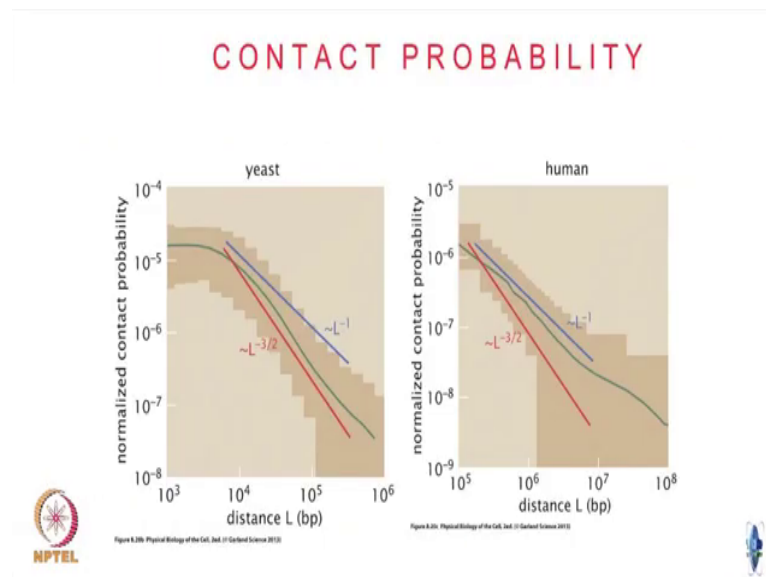
So, this is nice in that if; if I now go back and see what these experiments measured, it seems that if I looked at the yeast data, I could actually explain this yeast at least just by this measure, there are other measures which may or may not agree. But if I were to simply look at the contact probability and I would look at the yeast genome; it seems that what I would predict from a simple random walk calculation is actually what is absorbed for yeast, that it goes scales as L to the power of minus 3 halves.

On the other hand that is not true for humans. So, if you, you cannot use a simple random walk model to explain the contact probability of humans. And people have come up with various sorts of models in order to get this minus 1 exponent; I will not discuss them if you are interested in tell me later. So, that there is this question that what sort of a polymer model could you build, which would then explain this sort of a p c of S going roughly as S to the power of minus 1 and also ideally all of these other things that we have noted.

So, these are all things that whatever model that you build in order to explain this higher order organization of chromosomes, would need to sort of reproduce, ok. So, I said that different chromosomes might fold in different ways. So, is that true?

(Refer Slide Time: 17:27)
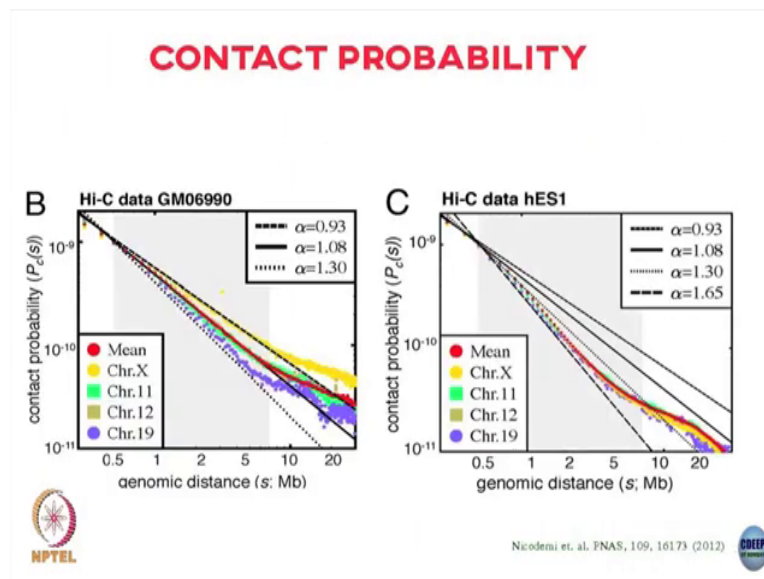
So, again let us look at these experiments. So, this is experiments on I forget which cell line this is, I think this is a human fibroblast cell and this is an human embryonic sensor. So, let us

look at the fibroblast cell first; the different lines are for different chromosomes. So, the chromosome x 11, 12, 19 and you plot the contact probability for each of these chromosomes. You will see that each chromosome lies slightly differently on a different line right, each has a different alpha; I was calling that beta, they are calling it alpha or whatever. So, this is the same as my beta over here.

So, each of them has a different exponent, they roughly lie around 1, so 0.93 to 1.3, but they are distinctly different; which means that, each chromosome has folded unto a different degree of compaction. If you were to do it in a different cell line, you would see different exponents for each of these chromosomes.

In particular if you look at human embryonic stem cells, it seems that this distinction has actually gone away. So, all these chromosomes now fall exactly on the same line, which is roughly what around which line is this point 1.65; the higher the exponent the more loosely it is packed. And so, remember what I said is that, when you packed something very tightly, it is more likely that this region does not contain any necessary genes for that cell type, right.

So, in the embryonic cell type, this is in the embryo, so it has not really how should I say; specialized to a particular cell type right, it is not differentiated. So, it could, this is my embryonic stem cell it could become whatever heart, liver whatever; but in this state it has not,.

(Refer Slide Time: 19:22)

So, all there is no distinction between what genes I need for functioning of a particular cell type and what genes I do not. And therefore, all of them roughly fold with or all of them roughly compress or package to the same extent; there is no distinction between the chromosomes. As you go through differentiation, as these embryonic cells divide and differentiate into these different different cell types, you will see this sort of it is starting to diverge from this behavior and ultimately in a completely differentiated cell type like this fiberglass, you will see different chromosomes folding completely to different extents.

So, these are dynamic. So, the this information is not there in the sequence, because the sequence is the same in this cell versus this cell; but these are dynamic. So, this is something apart from the sequence, apart from the genetic information which tells you basically how this sequence information is to be used by the cell in a functional way.

(Refer Slide Time: 20:34)
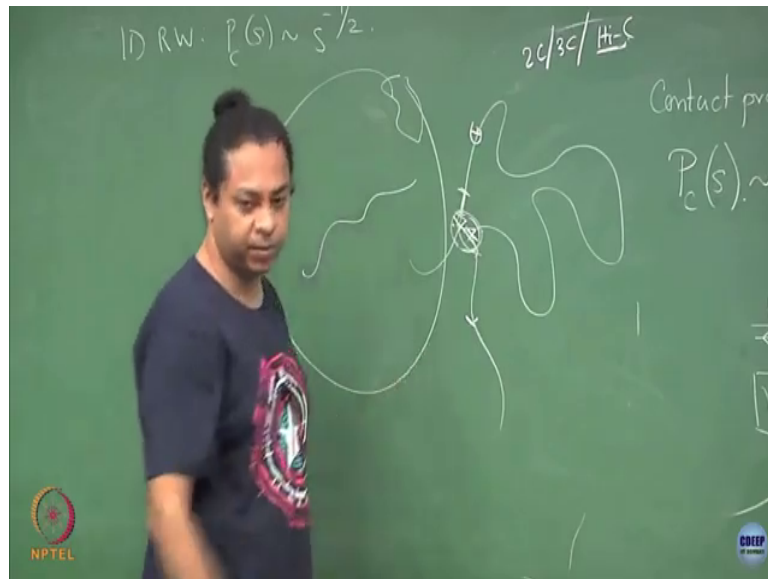
Moreover from these, so these are all obtained from these chromosome confirmation captures of experiments that I said. These are technically called various forms; so depending on the sophistication of the experiments, I let say these are called Hi-C experiments, ok.

So, these are all from these sort of experiments, but you can do even better. So, these are averaged remember, these are averaged over all chromosome, these are averaged over all base pairs which are S distance apart. So, I took this base pair and this base pair which was S apart, I found out what is the probability of them coming together; I took this and this which was also S apart, I found out what is the probability of them being in contact; I averaged over all such pairs in order to generate this contact probability curve. But I have more granular information than this which is that, how I often do exactly this pair and this pair come in contact and you can show this that information as well. What?

CHROMOSOME CONTACT MAPS

Chromosome 20   Chromosome 14
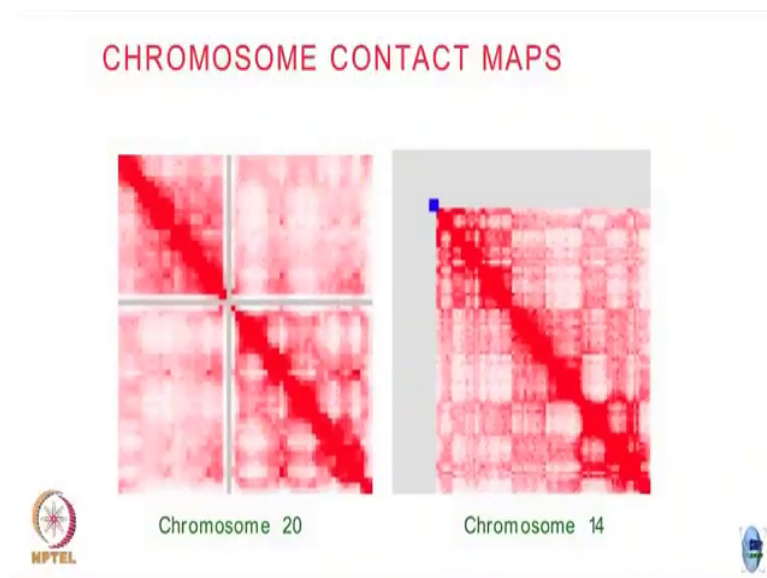
For example, because the figures are so bad, let me just draw it. So, both axis let say is some particular chromosome; what let say chromosome 20, chromosome 20, chromosome 20, ok. This is base pair 1, 2, 3, 4 all the way up to whatever number of bases the chromosome has, that particular chromosome has; the y axis is the same 1, 2, 3 all the way up to n, right.

So, if a chromosome has n base pairs, I can construct an n cross n matrix. And the color intensity tells me how often this pair has come in contact with this; this pair has come in contact with that and so on.
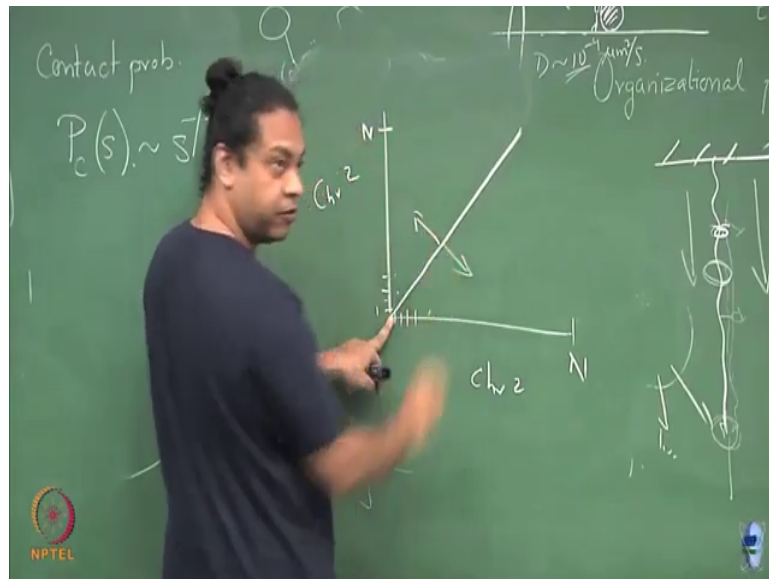
Ordinarily what would you expect? You would expect that well, there would be a sort of; it is drawn like this, but whatever since I written like this.

Ordinarily I would expect that the highest intensity would be along the diagonal right; because they are anyway close along the backbone itself. And then as I moved away from the backbone, this intensity would follow, right. What you see is that that is not strictly true, there are actually domains where even though they are far apart.

So, there are these domains like this where even though they are far apart, they have very strong interactions with each other. And these domains actually, we now know they have actually an important functional role to play. So, for example, if there is; if there is a protein which always brings these two into contact; then these this base pair and this base pair would

show a very bright red spot in these figures, and that tells us something about the structure of these chromosomes.
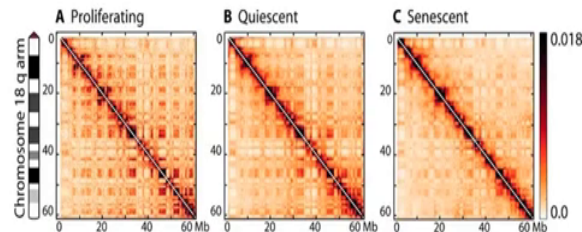
(Refer Slide Time: 23:21)

And you can do this for all different chromosomes; so for example, these are all the human chromosomes 1 to 22 and then chromosome ends, ok. So, it tells us how these chromosomes are organized, each individual chromosome is organized for a particular cell type; if it is this for another cell type, we would see some other sort of a feature emerging.

CHROMOSOME CONTACT MAPS

http://www.aidenlab.org/juicebox/

And also even in a particular cell type you can see that these features will change depending on the state of the cell. So, for example, if this is a prolife, this is the same cell, this is the chromosome. So, these this is called a contact map, this sort of a matrix is this matrix that I have constructed is called the contact map, the chromosome contact map.

So, if I plot this chromosome contact maps let say for chromosome 18 for the same cell type; but at different points in the cell in the life cycle of the cell. And you will see that the features that one sees are actually different depending on at what state of the life cycle is it in. So, not only does it depend; so this contact maps not only do they depend on the cell type, not only do they depend upon the cell type, but even in a single cell type it depends on time, this is what time relative to the cell cycle time of the cell lifetime as to what cycle it is in, ok.

So, these are very how should I say, these are very rich data that one has; now because of these extremely beautiful experiments that people have devised, the challenge is to sort of come up with a polymer sort of a model or whatever model that you want it will just take all of these disparate experiments and sort of tell us exactly what is going on, when this chromosome four goes inside the nucleus and organizes itself. Which is still an open question, we do not know; but at least whatever model that we come up with must answer or must agree simultaneously with all of these different experimental observations that we have.

By the way this website over here actually let us you this all of this data is open source. So, you can if you go to this program called juice box, you can upload whichever human chromosome that you want and you can see how these contact maps look for each different human chromosomes and you could average and you could generate those contact probabilities that we generate it. So, it is a good resource to play around if you are interested. So, I think I will stop here for the day, yeah something.

Student: Yeah.

Yeah, I will stop here for the day and I will continue with polymer modules in different context proteins and so on next class.