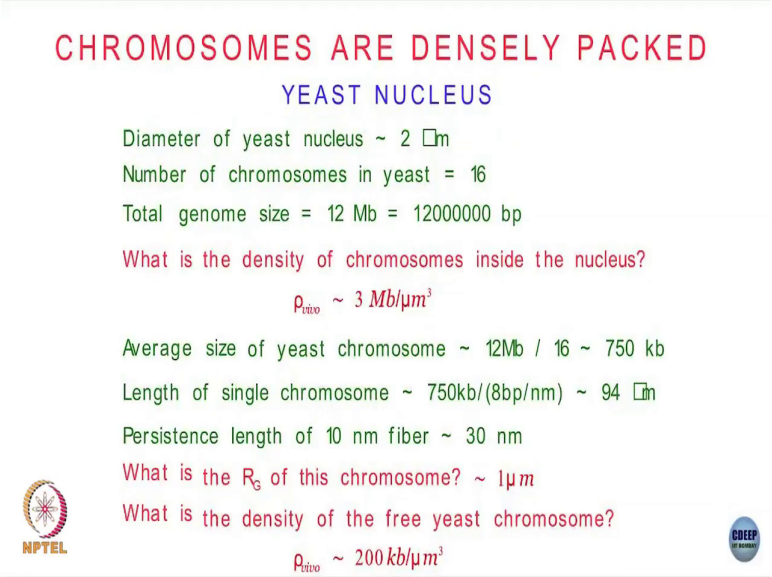
Physics of Biological Systems
Prof. Mithun Mitra
Department of Physics
Indian Institute of Technology, Bombay

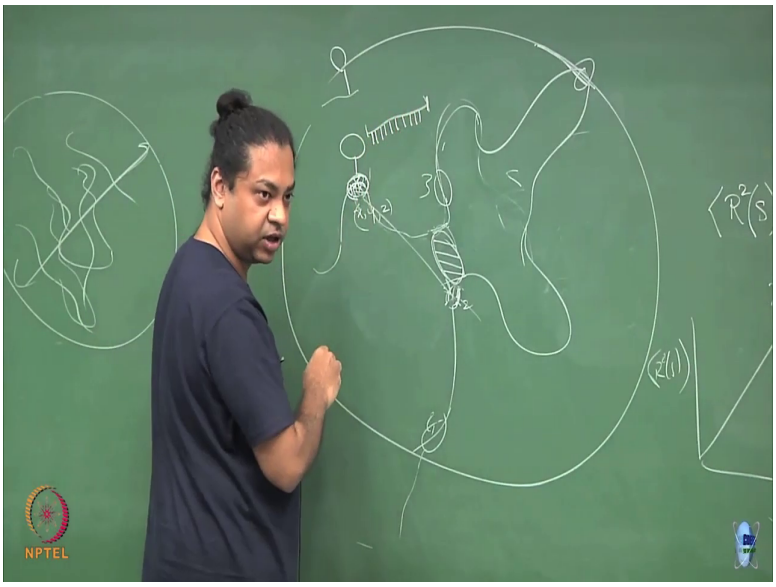
Lecture – 33
FISH and DNA Looping

(Refer Slide Time: 00:18)



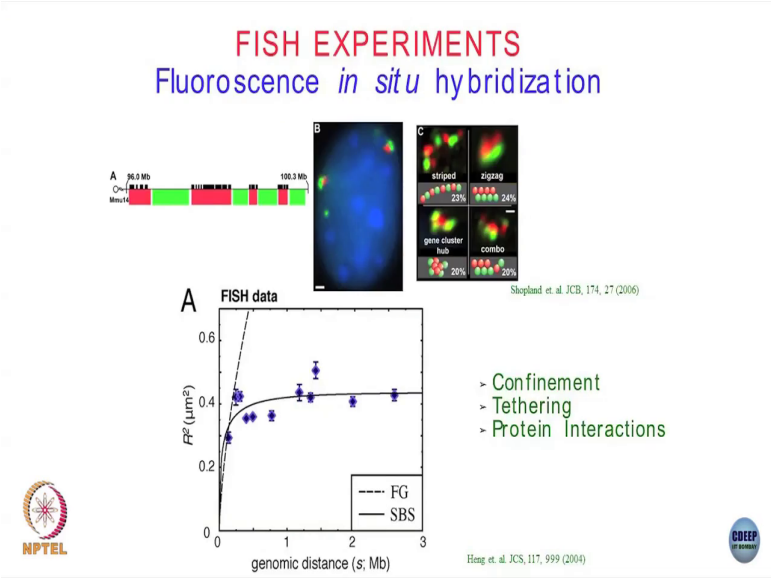
What else can I predict? So, people have done these experiments to see that, if I take a polymer let say I take this polymer, I take some random polymer and I look at two sides along this backbone, let us say one side here and another side there. If I walked along the backbone, these two sides are separated by whatever is the distance between them, let say some s base pairs, but I that is the genomic separation between these two base pairs.

(Refer Slide Time: 00:50)



But I could also ask what is the spatial separation between these two base pairs, what is the distance between these two base pairs? Right. And I could do, people have done experiments to find out these quantities; what they do is that, with this technique is called fluorescence in situ hybridization or FISH.

(Refer Slide Time: 01:11)



So, what they do is that; if you want to know where a particular sequence of this DNA is, you create a complementary is you create a sort of probe which is exactly the complementary sequence to this region, right. So, you create a complementary probe, so if this is ATTA this would be TAAT, right, and you attach a fluorescent marker to it. So, this probe will go and attach to this specific region of the DNA that you are interested in, right.

So, you can know that inside this nucleus where exactly is this segment of the DNA. You can do this for every segment; and therefore get, so if you tag this segment for example, this using a different target of course, depending on what the sequence is over here you would get where this sequence is on an average inside the nucleus. And then you could say therefore, what is the average separation between these two segments.

If you did it for a standard polymer, so it is just purely from a let say a random walk polymer; then again this separation between these two base pairs which was separated by S distance along the contour, would typically go as S to the power 2 nu, where nu is equal to half for a random walk polymer, R square would go as S. So, if nu is half, this is S to the power of 1. So, that is what I would expect for a typical polymer.

So, it would keep growing like this, so if I plotted R square S versus S, it would grow somewhere. You can do this, so if you do this experiment for actual genomes so here is a plot of that; this is actual fish data, this is the R square as a function of the average genomic separation between base pairs ok. So, this is averaged over what? This is average this is not for a particular pair of base pairs, so if these two base pairs are separated by S.

I take the average of these two over many many cells; but I take such average overall base pairs which are separated by S. So, this is one pair, 1 and 2; then this could be another pairs if I start somewhere over here, the other pair would be somewhere over here, right. These two would also be separated by S.

So, average over all base pairs which are separated by this S genomic distance and I find out what is the average separation between such base pairs. And, it turns out that yes it initially does increase; but then slowly after a point after around 1 mega base or so, it quickly approaches the saturation. So, it does not keep growing indefinitely as I would expect or as I would expect naively; it reaches some sort of saturation pretty quickly, ok.

And what could the reasons for this saturation be? A trivial reason of course is that, it is these polymer systems are confined, you are putting them inside a confined volume. So, they cannot really go beyond the confines of this nuclear radius, right. So, there is some sort of a confinement effect. There could be other effects like tethering; it is known that there are proteins on the surface of the nucleus, which I will discuss, which actually like to bind to these chromosomes and that again will restrict what sort of confirmations are allowed.

And, then there are also protein DNA interactions. So, if you have proteins which bind regions of DNA something like this; if a protein came and bound two regions of DNA like this, then these two base pairs over here and here, even though they are very widely apart on the backbone would be extremely close in spatial separations, right.

So, this sort of a leveling out of this R square could in principle be due to confinement, it could be due to tethering, it could be due to protein interactions or more likely due to all of these together. So we have, but whatever model you sort of build should reproduce this sort of a feature that it should sort of plateau out after around 1 mega bases, but 2 mega bases, ok.

Yes.

Ha.

Student: This paper is for (Refer Time: 05:34).

No. So, this is I think this paper was for a mouse genome. So, this is averaged over all sort of what do you mean a single DNA, right.

Student: Sir instead there yield (Refer Time: 05:57) come to that end also.

Right.

Student: That would not (Refer Time: 06:02).

No it did not. So, let say this and this. So, I have a fluorescent protein marker over here, which tells me the location x, y, z of this segment.

Student: (Refer Time: 06:17).

So, it is not a single base pair level. So, you build the target strand has, I forgotten maybe around I do not know I think it is around 30 to 50 base pairs. So, you build a complementary strand oh sorry ok. So, that was the confusion. So, it is not a single base pair, it has it is a target sequence basically, ok.

So, then that target sequence will come and seek it is complementary target sequence and bind. But in the level of this genome which is like 10 to the power of 9 base pairs, a 30 to 50 base pair pro is like almost a single point effectively, right; it is a very small point on this whole DNA backward.

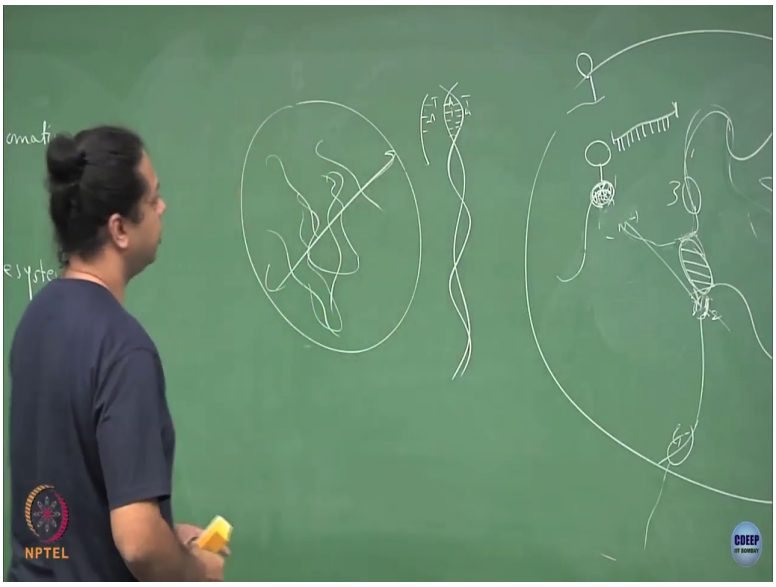
Student: Sir.

Yes.

Student: This (Refer Time: 7:05).

I see ok. So, you mean that. So, that is true.

(Refer Slide Time: 07:26)



So, you have this DNA double helix, right. So, which has let say A T whatever T A something like this and I am designing something which has anyway this complementary sequence this T A whatever. So, in order for it to come and actually bind here wait. So, one thing is that. So, I am not sure of the exact answer of this; but one thing is that, you therefore have two sort of possible binding partners, right. It could bind to this complementary strand or it could bind to this target strand right, with equal sort of energies if I have not done anything else, right.

So, even with even if I were to do nothing else; these would occasionally bind to this target strand and cause a fluorescence, maybe in only half the samples. But I am not sure that is the answer actually; I think they might do something so that the energy of attraction to this complementary strand is slightly lower so that it preferentially wants to bind to this. I am not sure if they do it by opening up the DNA somewhat in this region or some other method; but I

can look that up. But even if I were to do nothing I would still expect a sort of fluorescence signal, because I have introduce this target probe.

Student: (Refer Time: 09:06).

Yes.

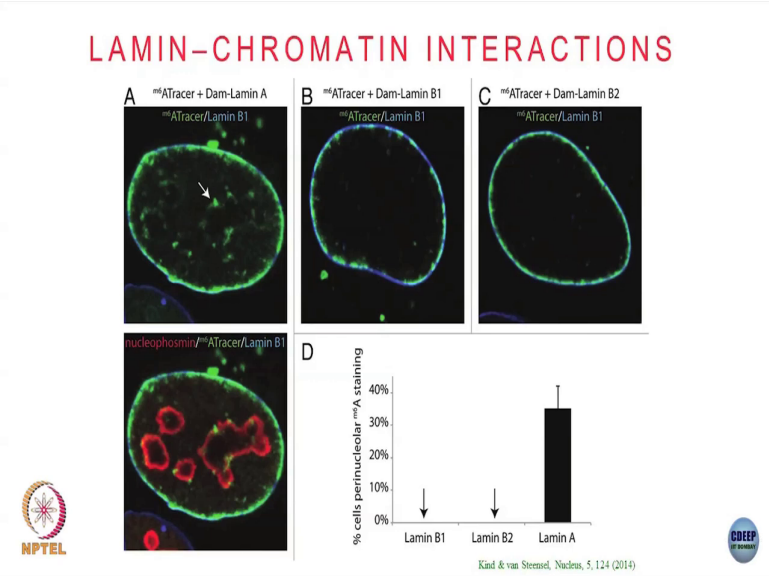
Student: If there is two copies of (Refer Time: 09:09).

Yes.

Student: There is (Refer Time: 09:12).

No right, you have two copies of every chromosome; when I say humans have 23 pairs of chromosomes, you have two copies of 1, two copies of 2 and then one x and one y. So, 22 pairs and then one x and one y and then when the cells divide you know this chromosome sort of. Let us come here, ok. So, let me just, the confinement is sort of obvious that; because you have a nuclear volume, you have some confinement effects.

(Refer Slide Time: 09:54)



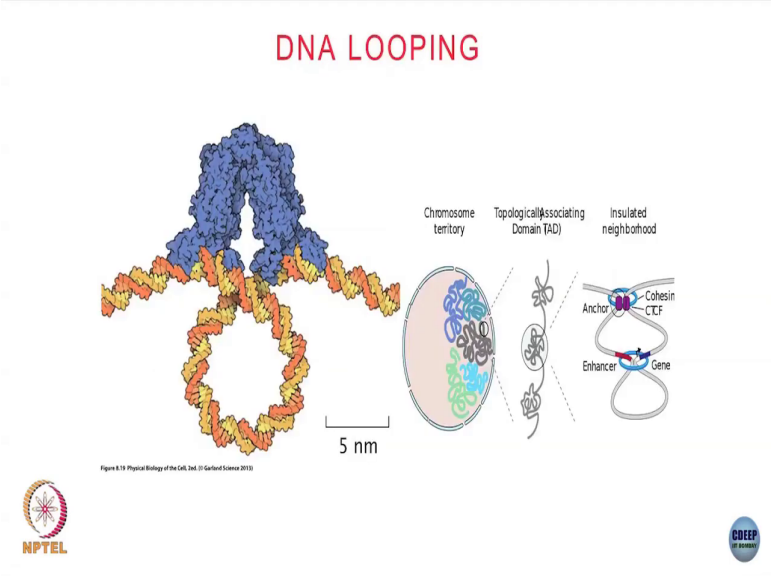
Let us look a little bit about this tethering proteins. So, here are particular class of proteins called lamin proteins and these lamin proteins are preferentially found along the periphery of the nucleus sorry. So, forget this word, just look at these things. So, there are different category, difference species of lamin proteins closely related; for example, this is lamin B1, this is lamin B2 whatever. But the key thing is that these lamin proteins are found sort of distributed along the periphery of the nucleus.

And people have done experiments to show that there are regions of the chromosome which bind very strongly to these lamin proteins on the periphery. So, you have these sort of tethering interactions which says that this polymer is not simply a free polymer inside this volume; but there are constraints which says that certain regions of the polymer will actually

be bound to this nuclear to this boundary which is the nuclear wall or the nuclear lamin as it is called. So, that is one class of interactions that can also provide some sort of compaction.

So, if you had a polymer which was interacting with the wall in this way; the statistics of that if you were to calculate average R square S would be different from the statistics of a free polymer, floating inside this bounded volume ok, so that is another thing. You have these tethering interactions.

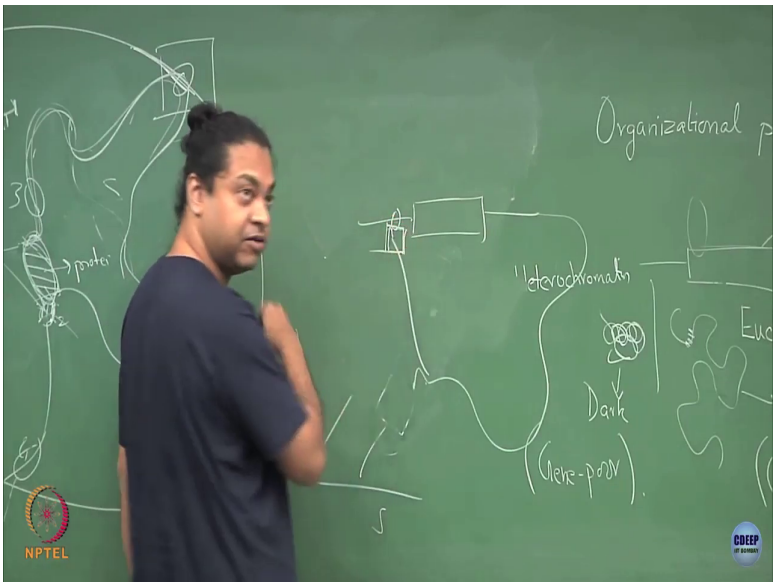
(Refer Slide Time: 11:15)



And thirdly you have these protein mediated interactions, which sort of comes under these very interesting thing called DNA looping. So, often people have found that there are these proteins which loop distance segments of DNA. So, this is some protein which takes together two segments of DNA which are very far apart and it sort of brings them in close proximity to one another.

So, what I have got now over here is what I would call a DNA loop. And people, these loops play a lot of functional role. So, for example, you could have a gene activator region. So, for example, let say I have gene, I have a gene sequence like this and in order for this gene to be active; then there is a promoter region which sits very far away let us say somewhere over here.

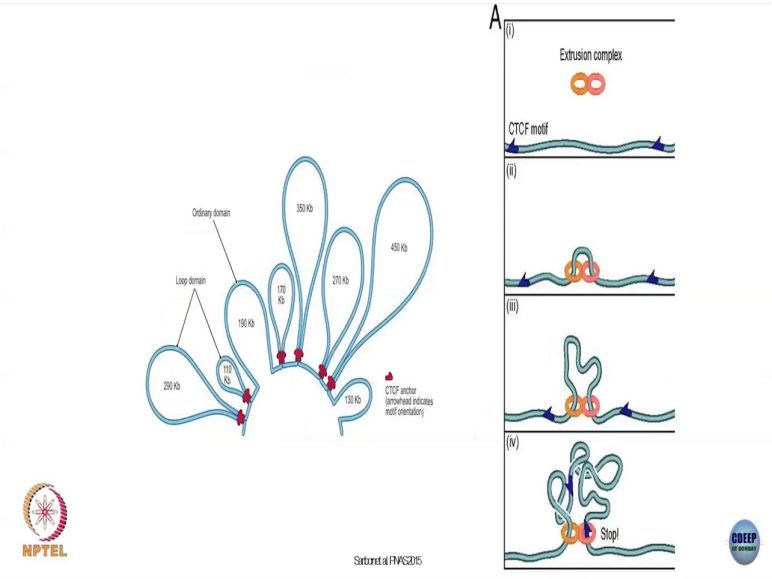
(Refer Slide Time: 12:09)



Far away along the backbone, if there was now a looping protein with sort of bound this region to that; so that my confirmation now looks like this, ok. Then this could be, then a protein could come and bind to this activated region and start transcription of the DNA or stop transcription of the DNA; need not be an enhancer it could be a repressor. But by bringing together distant segments of the DNA, you could play a role in regulating the genome itself.

So, the structure of, so the structure is important in order to sort of understand how the genomic information that is there in the DNA, itself is interpreted; and that is interpreted through this higher order sort of regulation. So, if you bring together distant path, some genes may start expressing, some genes may become silent. So, they say something about actually these looping proteins; it is actually in quite interesting or rather it is something that I am interested in, so I might as well tell.

(Refer Slide Time: 13:07)

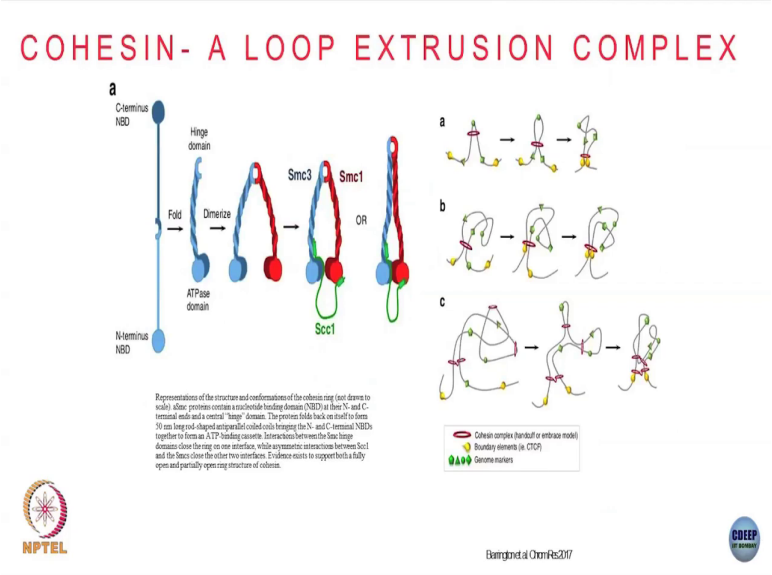


So, what is known is that you form these extremely large loops or rather loops at a all length scales; it is not just simple short loops that you would expect randomly, but you form extremely large loops of the order of kilo bases. So, hundreds of kilo bases even and in occasionally even mega base long loops and these are very well known in the literature, a lot of the functional roles of these loops are also known.

What was not know is that, what sort of proteins form these loops. So, what people hypothesized a few years back, 3, 4 years back is that; you have some sort of a protein which is call which I call an extrusion complex or which people call an extrusion complex, which has something like this structure it looks like 2 links, ok.

What it does is that, it comes and binds to this DNA topologically so that the DNAs; if I have two rings, you imagine a thread passing through both of these rings. And then as this proteins as this sort of DNA, I could pull on this loop or this thread and I could get longer and longer loop, ok. And it this loop formation would stop at certain specific sequence markers which are called CTCF motif. So, these are certain sequences on the DNA, which tells these proteins to sort of stop and that is where the loop stabilizes.

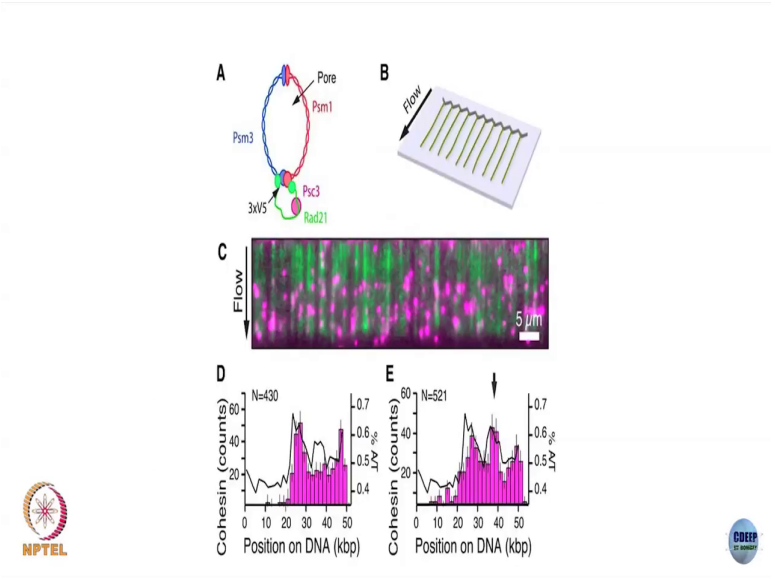
(Refer Slide Time: 14:37)



So, it was not really known what these proteins were; this was more of a hypothetical model. But recently a couple of years back people have found out; identified certain candidates for what are call this is loop extrusion complexes. So, these extrusion complexes of all the loop extrusion complexes; so for example, this is one such protein, it is a protein called cohesin which has a structure like this, so this is a dimer actually.

So, it could be that this DNA sort of one arm threads through here and the other arm threads through there and you form this sort of a loop. So, you sort of a cartoon, you form a loop any sort of thread through it and you keep growing the loop until you hit the markers at the ends, which tells the proteins when to stop.

(Refer Slide Time: 15:20)

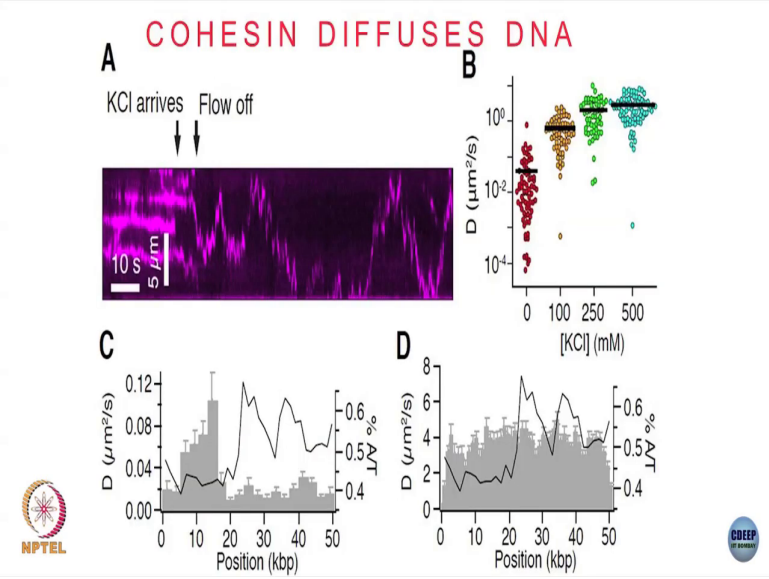


So, how do these proteins actually extrude loops and people have done these experiments as well, these are a nice set of experiments um. So, this is a DNA sort what is called as a DNA

curtain; what you take is that, you take DNA fibers you take a DNA strand for example, like this, take many many such DNA strands and you put these proteins on them, ok. And you put these proteins, these cohesin proteins on them and you tag these protein cohesin proteins with some fluorescent marker.

So, as these cohesins move about on the DNA, you can follow the trajectory of these markers and see what sort of a trajectory you get. So, these are these multiple DNA curtain, so this is called a DNA curtain. So, these are these multiple DNA strands. So, you could think of this like an ensemble ok; this is an experiment which is many many copies of this x simultaneously happening and then you can average over these ensembles to get average statistics. So, that is what we, forget about the lower figure.

(Refer Slide Time: 16:24)

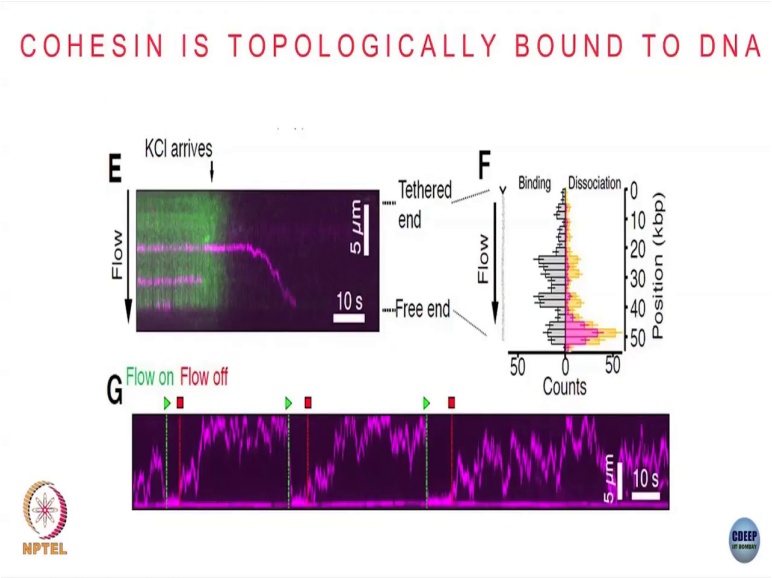


So, if you look at any such strand of; if you look at cohesin on any such strand and you look at the trajectory of cohesion; so now, we have to imagine that this axis is time, ok. This is axis is where my DNA strand is ok, so as I move along here; so these are continuous snapshots of that fluorescent marker. So, as I see like this, what I am seeing is the trajectory of that cohesin molecule on the backbone of the DNA. And you can calculate, so let say I start from somewhere and I can calculate how far I have gone in a certain amount of time, right.

So, I can calculate this by averaging over all these DNA strands and all times; I can calculate how much this cohesin moves on this DNA backbone as a function of time. And it turns out it moves perfectly diffusively. So, for diffusion this coefficient should be 1 and experimentally I think they determine this coefficient to be 0.98,, to which is pretty impressive, right. So, it is actually move diffusively; not only that you can measure therefore, the what is the diffusion coefficient.

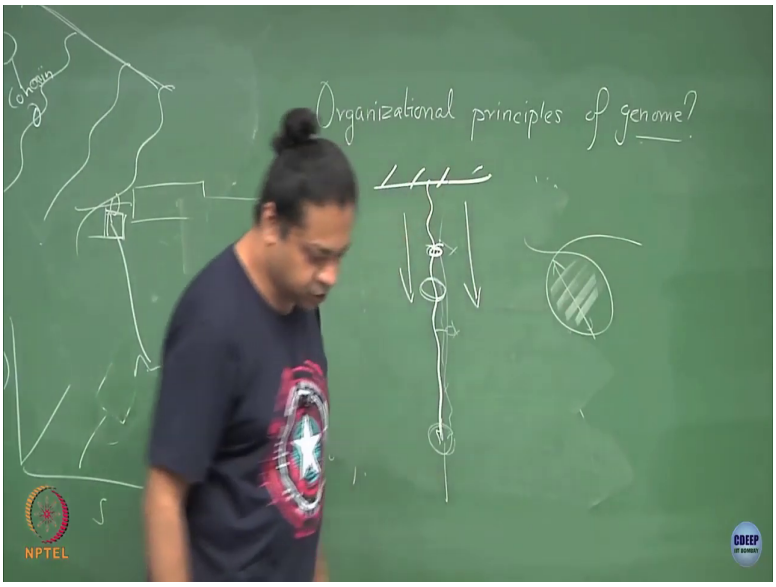
And they have done that experiment and they have measure the diffusion coefficient at different solid concentrations; physiological solid concentrations is roughly somewhere over here between 100 to 250, so inside the cell in a certain solid concentration. So, over there typically; so at physiological solid, the diffusion coefficient of this cohesin molecule which is forming this loop would be somewhere 0.1 to 1 sort of micron square per second. Again disregard lower frequency. What does it mean? So this further what you see is that, this cohesins are actually topologically bound to DNA, they are not chemically bound.

(Refer Slide Time: 18:11)



What does, how do I see that; so for example, I can look at locations where this binding and unbinding of the cohesin happens. So, you will see that binding, so here is my DNA strand; tell me if I am not being clear.

(Refer Slide Time: 18:46)



So, let say here is my wall to which I have attach my DNA strand and here is my cohesin protein which is doing a diffusive random walk on this DNA backbone. Now you could figure out where these cohesion proteins bind ok, at what region they bind. And you will see that; if you look at the histograms they bind sort of F. So, this is my position along the y axis is the position along the DNA and these histograms on your left are your binding events.

So, they bind sort of everywhere along the DNA, there is some sequence specificity but on an average they bind everywhere. On the other hand if I look at where they come off, they come off only at this end of the DNA, which is this free end of the DNA, which is floating around. So, it does not, once it is bound it does not unbind from random positions, but it sort of travels to the end and then slips out from here; which is consistent with this sort of a ring idea that one has, it is not a chemical bond but it is a sort of topological bond.

So, it forms a ring like this and then that ring does a diffusive work and then when only when it reaches the end it sort of falls off. You could show that for example, if you capped this end as well ok; if you cap this end of the DNA as well then this would actually not fall off. So, this flow is basically these are in done in nano fluidic channels, if you have some fluid sort of flowing to push the cohesin along this direction. So, when the flow is on, all the cohesins are at this end ok; they have all reach the end, but they cannot come out because they have stopped, I have put a cap over here.

And then when the I take the flow off, it goes back to doing a diffusive random walk; again I start the flow again everything gets pushed here and then again when I turn it off, it goes back to doing a random walk. So, it has extremely long lifetimes. So, it is not a chemical bond that binds and unbinds; it is actually a topological bond or at least some of them form a topological bond, which once it is bound stays on for a very long period of time, ok.

So it is, what is the Y axis? This is the DNA strand. So let say it 0 base here and then whatever 50 kilo bases at this end. So, you have to imagine the Y axis as the DNA strand and the X axis is time or it is the position of the cohesin molecule on the DNA strand therefore, ok. So, you could ask that well ok; this is how cohesin moves on DNA, it diffuses on DNA well it is nice. But we just saw that DNA inside cells is not just bare DNA; it comes in the form of these nucleosomes right, it comes in the form of these nucleosomes.

So, these are these histone complexes and the DNA is wrapped around that; which is a sort of large object, right. The histone protein is a protein octamer, so it is a large object. So, what happens when this diffusing protein this cohesin faces a large object on the on it is way. So, again they have done, these same set of experiments actually, have a reference, this reference over here in 2017.