Cell Biology: Cellular Organization, Division and Processes

Douglas Koshland

Professor, Molecular and Cell Biology, and Genetics, Genomics and Development

University of California-Berkeley

Lecture 25

The cohesin complex and its functions-The mysterious biological function of chromosome loops.

(refer time: 00:24)

Hi, this is Douglas Koshland from the University of California at Berkeley, and today I want to tell you about the mysterious biological function of chromosome loops. Today's lecture will not be something you can find in a textbook because this area of chromosome biology is very active and changing and I wanted to give you a little sense of a problem in biology, which is unresolved, but exciting.

(refer time: 01:05)

So, we start with the fact that segregating chromosomes acquire two essential structural features needed for their segregation. In interphase, the chromatin is diffuse. During S phase each DNA molecule of each chromosome is replicated to generate two copies, which are called sister chromatids. And these replicated copies are tethered together to generate sister chromatid cohesion. Once cells enter mitosis these sister chromatids undergo a process of condensation that leads to a highly compacted form of the sister chromatids, which are visualized in many micrographs of mitotic chromosomes.

(refer time: 02:01)

So, in 1990s one of the really sort of big questions in the field was, what are the factors that promote higher order chromosome structure that are needed for chromosome segregation in a eukaryotic cell. And my laboratory and others discovered that these processes of chromosome organization were carried out by a magical family or SMC family of protein complexes.

(refer time: 02:30)

So, SMC stands for structural maintenance of chromosomes and it turned out that pretty remarkably they were not just found in eukaryotes but they were also found in prokaryotes. So, these are in all forms of life that need to organize their chromosomes. Furthermore, what made them particularly intriguing was that they had an unusual structure; they were long ,50 nanometers in length due to long coil coils and they were flexible which allowed them to achieve multiple confirmations which we will see is important, but at this time it was very intriguing is the why were they so big and why were they so long and so flexible.

(refer time: 03:09)

So, one of the first SMC complexes that was discovered was cohesin, and its role is to tether together sister chromatids to generate sister chromatid cohesion that I just told you about. Cohesion is bound to the newly replicated sister chromatids as they are being formed during S phase such that by the end of DNA replication the two replicated molecules are cohesed along their length.

(refer time: 03:33)

So, what is the importance of sister chromatid cohesion? Well it promotes proper bipolar attachment by both promoting sterically constraining the two replicated chromosomes to attach to microtubules from opposite poles of the mitotic spindle. And once attached, it stabilizes these sort of bipolar attachments because the consequence of those bipolar attachments generates tension as the microtubules try to pull apart the sister chromatids. But they cannot because they are tethered together by cohesin.

(refer time: 04:10)

Clearly the importance of achieving this bipolar attachment is in fact to make sure that the newly replicated molecules segregate to opposite poles during mitosis such that upon cytokinesis each daughter cell receives one and only one copy of each of the initial chromosomes in the initial cell. And that is clearly important for the proper function of the cell and subsequent cell divisions.

(refer time: 04:42)

So, condensin is another SMC complex and its job is to tether DNA in regions within a sister chromatid to generate loops that compact the DNA. And these loops can be seen in micrographs of meiotic chromosomes, here shown are the lampbrush chromosomes of the salamander but you can also see these loops if you partially denature the mitotic chromosomes.

(refer time: 05:15)

So, why do you condense the chromosomes? Well, one is to allow the cell to move them around in mitosis without entangling with each other. And the second is that the chromosomes have to achieve a short enough length such that when they get moved to the poles, they are not the piece of DNA does not extend over the plane of cytokinesis, which would then cleave the DNA and this is indeed observed if the DNA only partially condenses.

So, this was all sort of our understanding of chromosome structure as the result of being able to visualize them in the microscope. However, a new method came to visualize chromosome structure and that method was a very nice method developed originally in Job Dekker's lab.

(refer time: 06:12)

It is called 3C- chromosome conformation capture. So, in this really ingenious method what happens is that you take a cell and in that cell the nucleosomes may be organized such that two particular nucleosomes are in close proximity to each other. You can capture that by adding a cross-linking reagent, and now you digest, you break the cells open, and you digest the chromatin with the restriction enzyme; the result of this is the two nucleosomes and the associated DNA that have been cross-linked, the DNA ends will be in close proximity and can be ligated together to generate a chimeric piece of DNA. With the DNA stitched together from the two nucleosomes; you then generate paired end reads by DNA sequencing and the result of that is you get the chromosome coordinates of the contacting nucleosomes.

(refer time: 07:10)

So, how do you interpret, how do you organize this information to make it interpreted in interesting ways? Well, what you do is, let us imagine a situation where you have unstructured chromatin you most likely will cross-link two neighbouring nucleosomes. And those two nucleosomes let us just say for example come from chromosome one at position 1000 and 1200. So, those are the two coordinates then you plot them on a graph and you will generate a contact map. And so, for those particular two nucleosomes, because their two coordinates are very close to each other they will lie close to the diagonal. In contrast if the chromatin forms a loop, then the two nucleosomes that are crosslinked from each to each other may be very far apart from each other let us say at position 1000 on chromosome 1 and 5000 on chromosome 1. So, those two coordinates 1000 and 5000, will plot off the diagonal as a spot and because these graphs are actually sort of a mirror, there is mirror symmetry, the same coordinates will appear as a second spot off the diagonal in a mirror symmetric position.

(refer time: 08:30)

So, once you have this technology, you apply it, in this case was applied to metazoan chromosomes and in the first surprise came that in interphase they observed loops 550 to 1000 kbs in size and also topologically associated domains, which essentially are like a loop except that we just scrunched the

looping DNA together. So, this was a surprise, we think of interphase chromosomes as sort of being diffused and disorganized but in fact they were actually organized into loops of distinct positions that we could detect on our contact map.

(refer time: 09:13)

Second surprise came was that the formation of these loops depended upon our SMC complexes; maybe not surprisingly it depended on condensin-like complexes, which what I have already told you were thought to tether together loops in cis for to give you the kind of loops that you need to compact chromosomes for mitosis but it turned out that cohesin also is used in some cells to generate these loops in cis. So, cohesin cannot only tether together the two replicated DNA molecules in trans, but also DNA sequences within a sister chromatid in cis.

(refer time: 09:58)

Third surprise and this is the biggest surprise was for years we just thought that these SMC complexes were sort of like a little napkin ring that sort of held together two pieces of DNA within, or between two DNA sequences. But what was discovered is that these Smc complexes are much more complex in their activity, that once bound the DNA they can begin to extrude a loop of DNA.

So, they are active molecules and this may explain there very complicated and long structure, they are long sort of long length and flexibility is that is the kind of features you might expect if you are going to generate a machine that can move something. And this discovery was first made in vitro, by the Haering and Dekker labs. And here we have a movie of that and what you are going to see here is we have a piece of DNA that has been tethered to a piece of glass.

And we have added in this case a condenser molecule that solution and you will see that spot of bright spot is where the condensin is binding and it starts to extrude a loop and the loop is actually sort of collapsed on itself. So, it will appear sort of as a line instead of an actual loop. But you can see that the loop gets bigger and bigger as the condensin moves down the DNA towards the two tethered ends.

And this was a remarkable in vitro activity, really stunned the field. And beautiful experiments done by David Rudner's laboratory showed that this in vitro activity was also happening inside the live cells, in this case using the kind of technology that you know we described for the chromatin capture to in fact visualize that these loops were happening in vivo along tens of kilobases of DNA.

(refer time: 11:52)

So, why would you have these kinds of looping activity, what would be the purpose of organizing the interface chromosomes in the loops? We understand why you do it mitosis, you want to compact them. Why would you do an interphase where you do not really need to compact the DNA? And the answer came with the idea that maybe they help control the proper gene expression. So, you can

imagine along the length of a chromosome, you have an enhancer and a promoter that are important for the expression of a gene in the heart and another enhancer and promoter they are important for the expression of a gene and the kidney, happen to be neighbouring each other. And it is very important that the proper enhancer talk the proper promoter. So, if for example enhancer two here inappropriately talk to promoter one, you might get the expression of a heart protein or a kidney protein in the heart, and that is something that could be very bad. And so, it was important to make these sort of loops at positions where such that you could you know organize the gene expression units into the proper way. And in mammalian cells there was a factor that helped the cohesin called CTCF, which is the DNA binding protein which seems to help generate stops to position where the loops will occur.

(refer time: 13:11)

So, we know these loops form but what is the evidence they actually are important for gene expression? So, it turns out that you can look in cell types and in particular, there was known to be a DNA inversion, which caused polydactyly in mice, which is an extra digit and when they looked at that by this chromosome capture technology it turns out that that inversion also changed the TAD and loop associated with that particular gene that is needed for preventing polydactyly.

So, that sort of said, that you need a proper loop to get the proper expression the gene to control the number of digits, at least correlated with an observation. We also saw that if you remove the function of CTCF or cohesin, these loop determining factors, that you can lose the TADs and loops throughout the genome of a human cell line and that would change the expression of a thousand genes.

So, again there was correlation between perturbing TAD and loop formation and gene expression and finally we could inactivate the dosage compensation complex, which was a condensin like complex. And that complex normally is important for specifically controlling gene expression of genes on the X chromosome. Loss of the DCC, caused the loss of TADs and caused the loss the proper loss of the dosage compensation. So, again there is a correlation between perturbing the gene products that cause loops and perturbing the actual TADs or loops that are formed on chromosomes and the change in gene expression.

(refer time: 15:00)

Further sort of an evolutionary argument, which fit this whole concept was the yeast was thought to be a very poor model for mitotic interphase chromosome loops and that is because here are the 16 mitotic chromosomes; in budding yeast they do not look very condensed. So, it does not look like you know yeast has in some sense a lot of looping activity. And indeed, you from first principle you would think well loops will not be needed in interphase either because budding yeast the enhancer the promoters are very close to each other. So, you can control proper gene expression just by proximity of enhancers and promoters to each other.

(refer time: 15:37)

Furthermore, if you do the direct experiment you can say well can we see loops in yeast by our chromosome conformation capture? And on the left, you see a mammalian cell and on the right you see a yeast cell. And you can see that while you can see the off-diagonal spots diagnostic of the loops in the mammalian cells, you cannot see them very well at all along the length of a yeast chromosome.

So, we have got a model, everything seems to fit together, but you know one of the very exciting things about active research is you get more information and that makes you realize that things are not quite what you thought.

(refer time: 16:16)

And so, let us go back through these observations and you can see that the first was well yes we had an inversion that causes the loss of TAD and polydactyly and mice but more careful experiments showed that the inversion the cause of polydactyly in this inversion was really because you were moving together two sequences that were only far apart close together. And it did not correlate with in fact the subsequent loss of the TAD. So, the TAD was just sort of an innocent bystander that got changed when the real effect was bringing together in close proximity an enhancer and a promoter that were not necessarily normally next to each other. The second was that while sure in those human cell lines we change the expression of a thousand genes, but loops and TADs are present throughout the human genome. So, there were seven thousand other genes whose loops and tabs were dramatically altered by the loss of CTCF and cohesin and yet their gene expression did not change. And finally, you know while it was true that inactivation of this condensin complex caused the loss of loops and TADs and also the loss of proper gene expression on the X chromosome in worms, when did the experiment actually where you altered the TADs and the loops by mutating the cis acting sequences, now what happens is that dosage compensation acts fine. So it implies that you need this condensin complex but it does not has to act to generate specific loops in order to change or to properly control the gene expression of the X chromosome in worms. Contrary to the whole concept that the purposes of these complexes to you know act at specific sites to control the interaction of particular enhancers with promoters.

So I bring this all up just to show you that you know when we get new observations we change our view of we have to change our models and have to rethink a little bit about what things might be happening.

(refer time: 18:26)

So with this in mind, we decided to reinvestigate the formation and function of chromatin loops in yeast. And we in this case was Lorenzo Costantino and Rebecca Lamothe, a talented postdoc and graduate student from my laboratory, who got together with Stanley Hsieh who was a postdoc in Xavier Darzacq's lab and we recently published a paper in eLife, which describes this research.

(refer time: 18:50)

So, what we did differently from what had been done before was, Stanley had invented a slightly altered version of this chromosome conformation capture technology. So, we cross-linked the cells but instead of digesting with the restriction enzymes, the chromatin was digested with micrococcal nucleus which generates a much smaller piece of DNA, gives much higher resolution of the contacting nucleosomes. And then you after you ligate and purify the DNA you do the paired end reads as we have done before.

(refer time: 19:30)

So, the outcome experiment is shown here and what you can see are human and yeast cells side by side. And the blue dots or the purple purplish circles indicate loops that are found closer to the axis which means they are relatively small loops that are found both in wild type human and yeast cells. And then what you can see is, we have a factor that we, in fact Peter's Lab originally identified, which is called Wapl/Wpl, which has been shown to prevent expansion of loops, of these positioned loops. And so, when you get rid of Wpl, bigger loops can be formed and they can be seen as spots that are further off the diagonal and those are circled in black. And you can see that we see that in both yeast and human cells when we remove Wpl we see the expansion of loops to new positions along the chromosome. So, what I would like to point out here is that if I took off the label at the top of these columns and asked you which is a yeast and which is a human cell you could not tell because the patterns are very identical.

Loop smaller loops along the length of the chromosome in humans and yeast; bigger loops when you get rid of this factor that inhibits loop expansion. And so, that similarity now turns the table on what we think about yeast. So, it now goes from being a poor model to understand chromosome loops to actually a very good model.

(refer time: 21:15)

And why is that? Well, we are just at the beginning to understand how the biochemical activity that people have identified in vitro translates into loop formation in vivo. And turns out that yeast are really cheap to grow, they have a small genome size, which makes the DNA sequencing cheap and the computational analysis really easy. And finally if we want to dissect all the features that control loop formation how active the complex the SMC complex is what makes it stop, why does it form let us say in particular parts of the cell cycle and not other, all those kinds of things will require mutations specifically inactivates specific aspects of the SMC complex or the cell, and yeast is ideal for that and we in fact already have lots of mutations that we can plug into to correlate what biochemical functions and how they translate into the vivo production of loops on chromosomes. So, we are very excited to use yeast as a model system to study chromosome looping.

(refer time: 22:22)

Now when you look at a contact map, if I were to show you some soon, they are usually pretty complex in which you can have a single spot plus some smaller spots, all in these diagonals etc or triangles that kind of off the length of the contact map. And these can be explained by suggesting for example in the cartoon on the left that the sequence at position one interacts with a sequence at position two to generate spot position three and position four.

So how can you have a sequence interacting with three different sequences further downstream of the chromosome? Well, that could happen if the chromosome actually folds into a quaternary structure. And so, when you are throwing in the cross-linker it can crosslink one to two one three or one to four, and this has been a common interpretation of what these complex chromatin contact maps might indicate.

(refer time: 23:33)

But you know one of the features of again the value of the yeast because of our very high resolution sequencing and quantitative measures we can get because of the abundance of the material is some new visions of how a complex contact map might be generated. And this comes in this case from information about knowing where cohesin binds along the length of chromosomes using this technique called chromatin immunoprecipitation, in which the size of the peak reflects how much/ the percentage of cells, which have cohesin bound at that site. And this technology was used many years ago to define where cohesin bound on chromosomes. In fact, Shikha Laloraya, former postdoc of mine and a faculty in India was one of the pioneers to first use this technology to map cohesin's binding to chromosomes in yeast and subsequently by others it was used in mammalian cells.

So, we show those peaks of cohesin binding and what you can see is that the off-diagonal spot arises from the cohesin bound for example the first one at position two and bringing together the sequence associated with cohesin and bound at position three. Now one of the things we knew about the cohesin binding at these sites was it was stably bound.

So, the cohesin bound at the so-called cohesin associated regions or *CARs* is stable, but we also knew there was a second population of cohesin that was dynamic, that was binding and coming on and off chromosomes. And we assumed that that was likely the looping cohesin. And the third thing we knew in this study was that or we revealed in our study in eLife was that stably bound cohesin in *CARs* occurs prior to loop formation.

So, with these observations we came up with a model, which is you first lay down the stably bound cohesin, which is probably doing the cohesion, and then this dynamic cohesin gets on and starts pumping out loops.

(refer time: 25:45)

And it goes along until it runs into this stably bound cohesin and the activity stops. And that would explain how you could generate a specific loop at specific positions. Now, we have to incorporate

that model with the fact that the peaks of cohesin binding are not the same along the length of a chromosome.

(refer time: 26:13)

And so, we interpret that to mean that at some sites in a chromosome, cohesin is bound on that DNA in every cell in the population. While at other sites shown on the left here, the little red dumbbells, cohesin is bound at that particular site in only a subfraction of the total cells. And you can see if you sort of place those kinds of stops along the length of the chromosomes in different cells with different patterns, you are going to get different looping patterns.

So, each individual cell if we could do a contact map on individual cell, would actually have a different contact map. And if you take all these contact maps of the individual cells and mix them together as a population, which is like the actual experiment we are doing here, you can generate a contact map, which is identical to the to the one that is previously or often interpreted as being reflecting some specific tertiary folded structure.

So, this implies that the folding is much more dynamic in individual cells, and the pattern of loops is is different between cells.

(refer time: 27:25)

So, now these observations lead us to the idea that there must be some additional conserved biological function other than controlling gene expression. So, first of all as I said, we have these genomic loops that exist genome-wide in an organism like yeast where there is no reason to invoke a function in gene expression because it is not really needed, the promoter enhancers are close together.

Second, there is heterogeneous loop formation in individual cells in yeast, when you want to try to accomplish homogeneous gene expression. So, you know, how are you going to get all cells to- if they are- if all cells are expressing a particular gene but only a subset of them have loops, does not make a lot of sense if that is what the loop is doing is controlling or ensuring the proper expression that gene.

And third, you know in human cells as I explained earlier, there are loops all throughout the genome and yet only a subset of genes are causing changes in gene expression. So, what are the other 7000 loops that are existing that are not causing gene expression, what are they doing?

(refer time: 28:32)

So, we came up with this idea that maybe it is involved in DNA repair. So, imagine that you have two paired sister chromatids through cohesion and you get a double strand break in one of those chromatids in a repetitive sequence, labelled one and two here. If the break in one repairs off the

sort of repeat out of the identical position on the other sister chromatid, all will be good, you will just regenerate the two initial sister chromatids, as expected.

In contrast, if the break in one repairs off a repeat that is further down along the chromosome arm that can generate, depending upon how the repair happens, duplications and deletions, and that can be bad for lots of reasons.

(refer time: 29:21)

So, thinking about that in our model we can think about how looping activity might actually, by generating loops along the two sister chromatids, bring the broken piece of DNA in contact with the proper repeat, as opposed to the improper repeat. And therefore, directing repair of the broken DNA off the repeat that's not going to change the DNA sequence and restore the proper chromosome structure.

(refer time: 29:52)

So, that is our model at the moment, may not be right, but I think the bigger question of what loops they are doing in chromosomes remains mysterious. And I want to say that positioned loops is a conserved feature of all eukaryotes, and in fact, there are positioned loops in prokaryotes as well. And the SMC complexes act to actively form loops in all eukaryotes. So, this concept of motors is like a real wild factor for us in the field. The position of loops in the genome differ by the presence of stops; and there is a slight difference so, in mammalian cells we think the stops are made by this factor called CTCF, while in yeast it may be the stably bound cohesin. Although some recent experiments with mammalian cells that stably bound cohesin in the mammalian cells may be a second way of generating stops for loop formation. The existence of positioned loops throughout the yeast genome and in fact the human genome, indicates that they likely perform a universal function distinct from controlling gene expression.

So, I hope you will sort of appreciate this is a really active and interesting area of research and stay tuned to understand when the mystery gets solved in our little detective story some years down the line about what chromosome loops are really doing and how they are forming.

(refer time: 31:23)

And so, with that I want to thank again everyone in my lab who helped us and allowed this research to happen, particularly Rebecca and Lorenzo, and thank you for your attention.