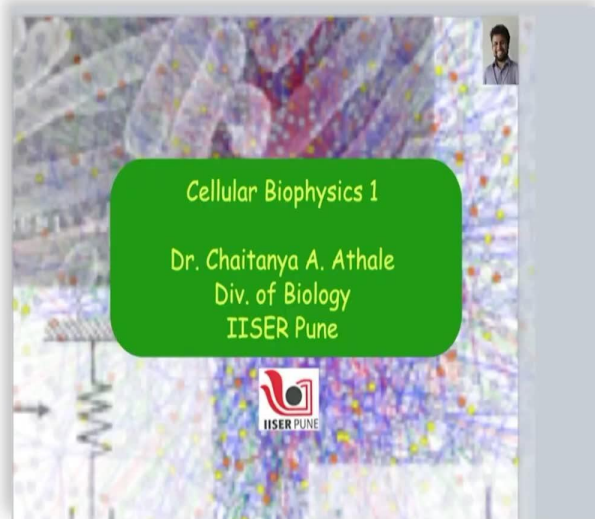**Cellular Biophysics**
Professor Dr. Chaitanya Athale
Department of Biology
Indian Institute of Science Education and Research, Pune
Understanding chromosomes as statistical polymers Part: 1

(Refer Slide Time: 0:24)

Hi, welcome back. I am going to continue from where we last left off and discuss what goes ahead of what we talked about last time, namely the kinetic gas theory in average molecular velocities, if you remember, we calculated these for lysozyme. We then discussed that if molecules cannot actually move like bullets in a ballistic fashion, then they are more likely to be

performing some kind of collisions, that there is no independence of the individual molecule walking, and that it might describe something called a drunkards walk or a random walk.

And we discussed the statistics of the random walk because some very nice properties come without assuming anything physical, the point of raising random, drunkards walking, besides the entertainment value, is that the statistical properties are independent of the underlying phenomena and this is a beautiful process because you can see that we did not have to say anything about whether it is an old drunkard, a male drunkard, a female drunkard, a non binary drunkard, child drunkard, grandpa drunkard, it does not matter whether it is molecule of polymer, we then ended with saying that if the step size average is not, not one, but in fact, Kuhn's length A, then the mean squared, this end to end distance, position at the end, becomes equal to NA^2.

Because it was in fact, we ignored that one in the simple arithmetic. And I ended with a question about rubber elasticity, which is due for Thursday, I would like to come back to that in a while. So, for today, I want to talk to you about random of polymers and probabilistic arrangements, average distributions of polymers, persistence length, and relating coolant persistence length to each other, ending with radius of gyration, which is again a part of your assignment for next week and I hope you are eagerly looking at it at least you had a look at it, if you have questions, Thursday is the right time for you to discuss it with me.

If there are any clarifications, obviously, I cannot give you the answers, but I can help you get to them.

So, let us go ahead. So, I mentioned to you that rubber elasticity has an unusual property and this is sort of seen in this diagrammatic view, that it consists of many hybrids or polymers in other words, which can be either stretched to extend them or when they are released, they retract and they become relaxed again, which in the case of polymers is a more compact form. So, we are trying to say that as the stretch is released, those the force stretch is released, the retraction actually results in an entropy increase. And that is related to the idea that when you stretch, we expect the polymer to warm up. And when you release it, retract it, we expected to cooled out for Thursday, I would like all of you to please bring a rubber band, so that we can test this out.

The idea is that there is something happening to the temperature of the material when we stretch it. And when we relax it and that is exactly the part of my question to you, for a written reply as an assignment. Indeed, metal springs are considered to be enthalpic machines and rubber bands are considered to be entropic machines or entropic strings. And this hopefully gives you the hint to the answer, but without hopefully revealing the whole answer.
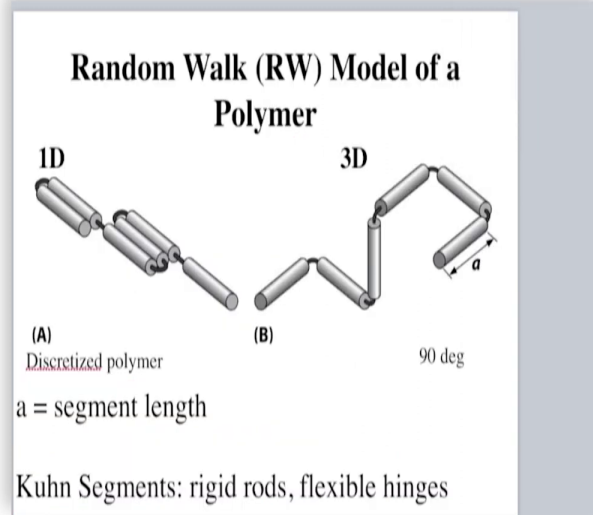
So, indeed, what is the special property that makes random walk polymers so unusual? I said this last time also that if we consider a one dimensional or even a three dimensional discretized, polymer meaning to say we consider a little rod segments, which let us say in one dimension, for simplicity, they flip one way, choose another segment, clip another way, and so on and so forth independently, flipping a coin effectively, where you either left or right random walk and in 3D you have more degrees of freedom.

These rigid segments then we call kuhn segments, which are inflexible in the sense that they are of a fixed length, and they do not bend or turn but they are connected by flexible hinges and this is why this model is also called freely jointed chain. It is like your necklace, I mean, some of you have these beautiful necklaces with segments, the segments may be little beads or cylinders, and the cylinders have little complete flexible joints. Here for simplicity, we assume only 90 degree swaps in 3D.

(Refer Slide Time: 5:33)

RW Polymer Configurations in 1D

N=3

n_r=no. of steps to the right

n_l=no. of steps to the left

p_r

P_l

For right steps:

$$W(n_r; N) = \frac{N!}{n_r!(N - n_r)!}$$

So if we consider now coming back to the one dimensional case, a chain of really joint chain of 3 segments, then we have a whole bunch of possibilities that we can arrange this polymer in. So, we can either have them as taking 3 steps to the right, 3 steps to the left, 1 step to the right, 1 step to the left 1 step to the right, so 2 steps to the right, 1 steps to the left. But in that sequence, then we have 1 step to the left 2 steps to the right. So still 2 steps to the right, but slightly different arrangement, as you see. 2 steps to the right and 1 step to the left, and so on and so forth. So, if we are to find out the multiplicity that arises, the possible confirmations that arise out of a certain number of steps to the right, then we have a very easy recourse to going back to our binomial coefficient to count the number of ways we can arrange it.

So, let us ask the first question, when we take the top case here, when you take 3 steps to the right, how many ways are there of arranging it? When you take 3 steps to the left, how many ways are there of arranging it? When we take 1 step to the right, 1 step to the left, 1 step to the right, how many ways are there have arranged it and so on and so forth.

(Refer Slide Time: 7:14)

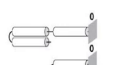And as you will see, the answer to that question comes as 1 for all right steps or all left steps, and the multiplicity increases when they are not just all 3 steps are not right or left, necessarily. I this is again, intuitive, but I would like you to ponder on it.

(Refer Slide Time: 7:39)

So, the polymer end to end probability that is to say, what is the density distribution after all these steps are taken, you know when what we talked about initially was just the discretized version, we took steps left took steps right we kind of did the calculation. But what if you

calculated average, if you do this 100,000 times there are hundreds of segments rounds of polymers, what will the probability overall departure from the origin B. In that case, we come back to the binomial coefficient times $1/2^n$.

And that is solved or resolved by arriving at an approximate Gaussian distribution of the probability or end to end distance, I want you to stare at this expression down below, where P of end to end distance R for N steps, meaning N segments. So, we have gone from drunkards to rigid rods kuhn segments is equal to 1 upon the segment length a into the under root term $\sqrt{2\pi N}$, N being the number of steps or segments and all this product of exponential in negative half r^2 end to end distance upon NA^2, you can see this here.

$$P(R; N) = \frac{1}{a\sqrt{2\pi N}} \cdot e^{-\frac{1}{2}\frac{R^2}{Na^2}}$$

(Refer Slide Time: 9:01)
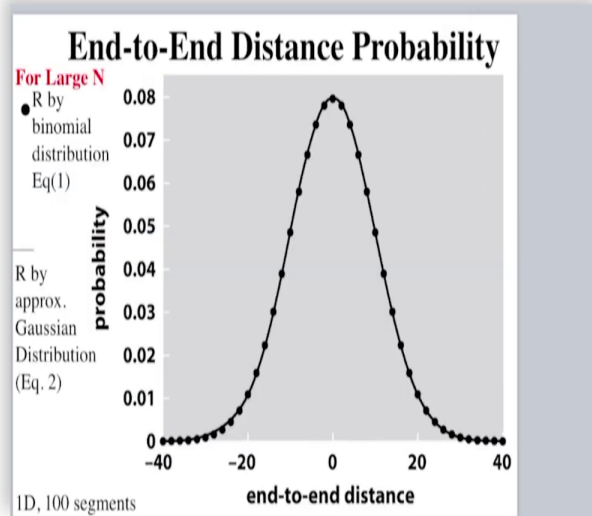
Probability for End-to-End Distance

Gaussian distribution

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Central limit theorem

Freely Jointed Chain

$$P(R,N) = \frac{1}{a\sqrt{2\pi N}} e^{-\frac{1}{2}\left(\frac{R}{a\sqrt{N}}\right)^2}$$

Now, this is indeed a Gaussian and indeed, if you plug this for a 100 segment polymer, you exactly get a Gaussian and you can, in fact, compare it to the standard Gaussian distribution, which you hopefully done in biostatistics or high school or I do not know where, it in fact tells you that the term a that we have been referring to a segment length can be easily mapped to sigma and N in that case, disappears in the Gaussian distribution. The numerator with the exponential term has the equivalent of sigma is equal to $a\sqrt{N}$ and that is how that N disappeared by the way.

And μ or $x - \mu$ is R it is the mean displace distance from initial position, assuming x is 0 all of that squared. This by the way, that Gaussian distribution comes from the central limit theorem, which I suspect you have had to study at some point. And because of that I am not going to elaborate on it but I am going to ask you, when we meet on Thursday, what is the central limit theorem.

So, in 3 dimensions, this was 1 dimension, obviously, this equation becomes a little more elaborate

$$P(R; N) = \left(\frac{3}{2\pi Na^a}\right)^{\frac{3}{2}} e^{\frac{-3}{2}\left(\frac{R}{a\sqrt{N}}\right)^2}$$

it is $3/2\pi Na^2$ so because of that Na^2, we have to take 1 by 2 in the power 3 to the power 3 because of 3 dimensions.
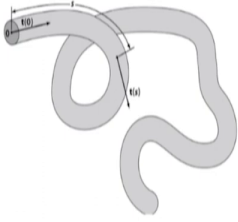
End-to-End Distance Probability
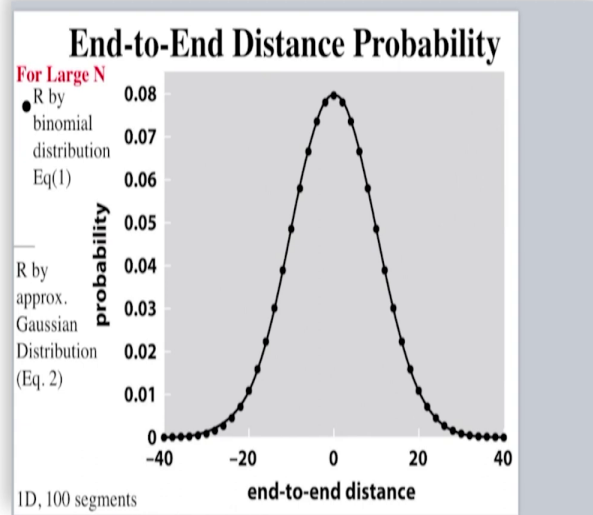
The interesting part of this is that the highest probability, as you also saw in this graphic for 1 dimensional system, the highest probability highest density of finding end to end distances is 0. So, and this is true of a random walk, also, which is another way of saying that the drunkard is not going to get very far. Now, they will most likely position that drunkard will be at his the lamppost or the fibers, the polymers in a rubber are most likely to be compact, this is a very strange way of looking at things.

Because in our normal spring mass system in the spring, the reason why we say they are going to reach a certain position is because the spring, atomic springs have a certain resting length. And that is the energy that we described using Lennard Jones potential of inter atomic interactions. Here, this is purely statistical and this is the beauty of it. And this is why I had spent so much time talking also about the osmotic engin but let us go ahead, not backwards.

(Refer Slide Time: 12:06)

Persistence Length (ξ_p)

- ξ_p: Length over which polymer essentially straight
- Local rigidity
- Tangents to polymer t(s1), t(s2),... at distances s1 and s2
- Freely jointed chain model

So, we now come to a term, which we will refer to as persistence length and we have talked about this earlier, I am going to do a slightly more detailed effort at trying to get at it. So, persistence length by definition is supposed to be the length over which a polymer is essentially straight. And it considers local rigidity that tangents to the polymer T of s1, T of s2, at distances s1 and s2 in a freely jointed chain.
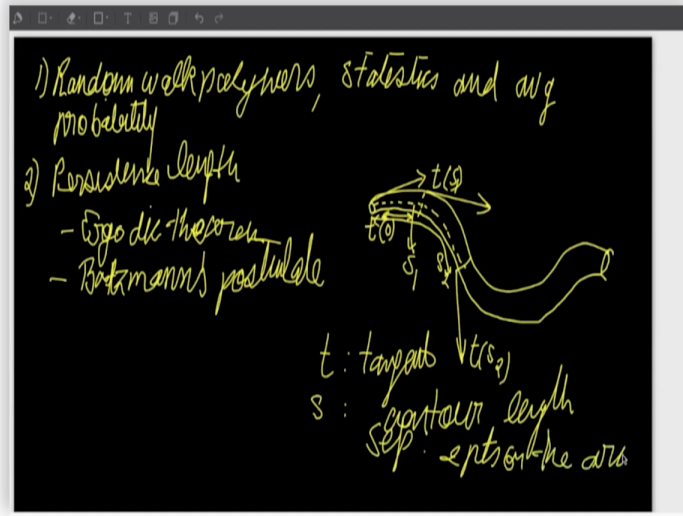
So, let us see what we can do about it. So, we just talked about random walk polymers, statistics and average probability distributions and we want to continue with persistence length. And in order to do that, we want to maybe discuss very briefly the so called ergodic theorem, which tells us that nearly every microstate approaches the arbitrarily close to the all of the true states of the corresponding ensemble in the course of time and a system that follows this therefore, is then following the ergodic theorem.

And I encourage you to look at your physical chemistry and statistical mechanics textbooks to refer to it again. And the second thing we also rely on for kind of getting at the persistence length is the Boltzmann's postulate the time average of faults and postulate that the time average of an isolated system is the average over all states in the microcanonical ensemble. And as I sort of mentioned in the slide, the persistence length of a system of a polymer in this case is based on taking tangents to the polymer along its length with increasing distance between the tangents or the tie points at which the tensions are taken.

So, if we call this tangent t at 0, then we can call this tangent at t at s, where s is nothing but the contour length between the 2 positions and this can be s1, s2, and so on and so forth where the contour length becomes s2, which is of course, the whole length from here to here. Whereas, this one is just s1. So, we define t as the tangent and s as the control distance or controlling separating 2 points on an arc, this arc is nothing but the one that defines the polymer.
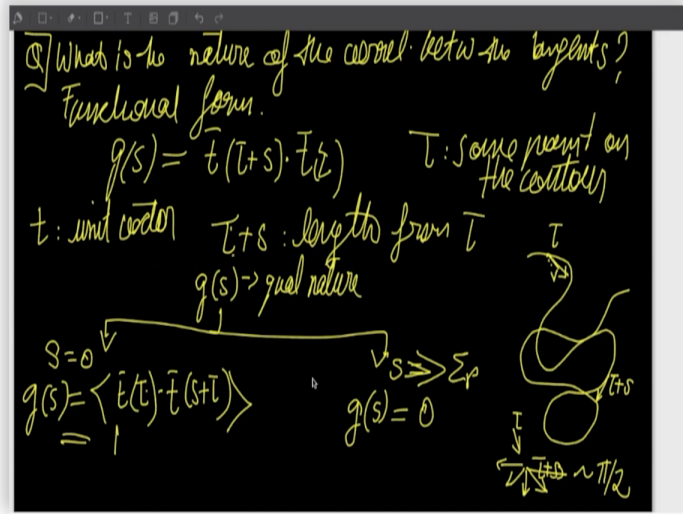
In such a case, the question we ask ourselves is, what is the nature of the tangent tangent correlation and what is its functional form, nature of the correlation between the tangents, in other words functional form. Now, some of you already know this perhaps, and either way we are going to guess and the guess we are going to make is, if we are going to say well, we took the tangent at some initial position $\tau + s$ and we tried to find its correlation its product with t at τ, where τ is just some point on the contour.

Then, indeed the tangent is a unit vector and $\tau + s$ is the length of the curve from τ so that we have a nonzero start point. In such a case, if you want to infer what this g(s) becomes, then we have possibly two cases that we want to consider. One case is that s is equal to 0 and the other that s is equal to s is much much greater than the region over which the polymer is straight, in other words, it is very winding, so, we are considering a very winding polymer, then if this was τ and s was very very long, then we would probably be referring to something like this much much greater than the region over which the polymer is straight.

And when s is 0, then it is essentially like taking the tangent at the same position I have drawn it in slightly different directions just for clarity.

$$g(s) = \langle \bar{t}(\tau) \cdot \bar{t}(\tau + s) \rangle$$

This would imply that g(s) which is nothing but the dot product of two vectors averaged I sometimes have been skipping the vector signs, but bear in mind I did not that was not intentional. So, $t\hat{}$ or $\bar{t}$ of τ dot product with the $t\hat{}$ of (s +τ) or (τ + s) averaged for angle of 0 is nothing but the product of the 2 magnitudes of the vectors times $cos\theta$, $cos\theta$ of 0 is 1 therefore, it is nothing but 1.

In the case that the two angles the two angles are not correlated they are not 0, then they can be all kinds of angles. So, essentially these two vectors at τcan be be τ and τ + $s$can take any value various angles. I am assuming 180 degree mirroring that would mean that on an average, I have about $\frac{pi}{2}$ and then that makes my $cos\theta$ cos $cos\frac{pi}{2}$ which is nothing but 0, 0 multiplied by everything 0. So, this is the two the nature of this g(s) qualitative nature. So, the question becomes what function satisfies this condition that when s is 0, the value is 1 and s is much much larger and it decays it is 0.

(Refer Slide Time: 19:50)
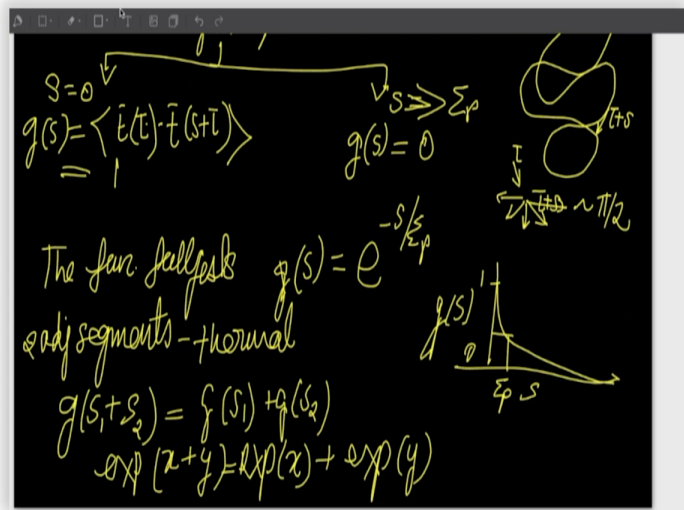
And the function that fulfils these requirements is surprising or not very surprisingly nothing but the exponential where we say that the coefficient of the exponent, the decay, exponential decay is in fact, the persistence length it is like saying that the correlation, like we said goes from 1 to 0 for increasing s, such that it is a continuous function with a characteristic decay length of persistence length. And that is in a way our idea of how we can intuitively understand the

meaning of the exponential. There are more formal derivations, and I am not going to do those. The second as maybe line of evidence is that when there are two adjacent segments, then and they are buffeted by thermal effects that is diffusion, then the correlation function g(s1) + g(s2) should effectively due to the fact that they are independent of each other be the sum of s1 and s2 and this property is also satisfied by the exponential where

$$exp(X + Y)exp(X) + exp(Y)$$

exponential of X plus Y can be written indeed as exponential of X plus exponential of Y sum of exponentials.

(Refer Slide Time: 21:59)
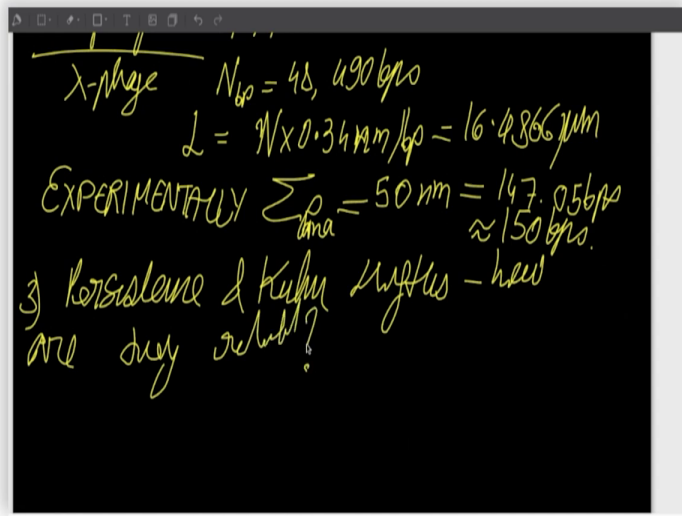
So, all this being as it is we can indeed now write the tangent tangent correlation

$$\langle t^\wedge(s_1) \cdot t^\wedge(s_2) \rangle = e^{\frac{-|s_1 - s_2|}{\xi_p}}$$

as the product the dot product of the two tangents $t^\wedge(s1)$ times $t^\wedge(s2)$ averaged as e to the power minus magnitude of the distance between these two segments divided by Cp at some constant which indeed is the characteristic length. So, we can summarize say that when L is equal to when s is equal to L, then indeed g(s) is 0. And when s is equal to 0, then g(s) is equal to 1. In other words, when the segments that we are looking at the contours are the identical contour, then that correlation term is 1.

Just to give you a sense of the order of magnitude and I may have discussed this in the bending rigidity measurements, but let us take a biopolymer that we all are very familiar with DNA sets it always catches everyone's imagination and we take it from a virus the lambda phage then, the number of bases of the lambda phage genome is 48,490 base pairs. The contour length, L is number of base pairs mutiplies 0.34 nanometers per base pair which turns out to be 16.4866 micrometres.

By experiments it has been found that the persistence length of DNA is equal to 50 nanometers which in base pairs terms is 147.05 base pairs which we can approximate to 150 base pairs. In other words 150 base pairs of DNA are effectively straight are persistent. And this is where we

want to perhaps relate the 2 terms that we have now brought up. One is indeed what I have been mentioning right now which is persistence length and the other is the one we started with that is Kuhn's length. So, the question is persistence and Kuhn's length, how are they related?

(Refer Slide Time: 25:51)
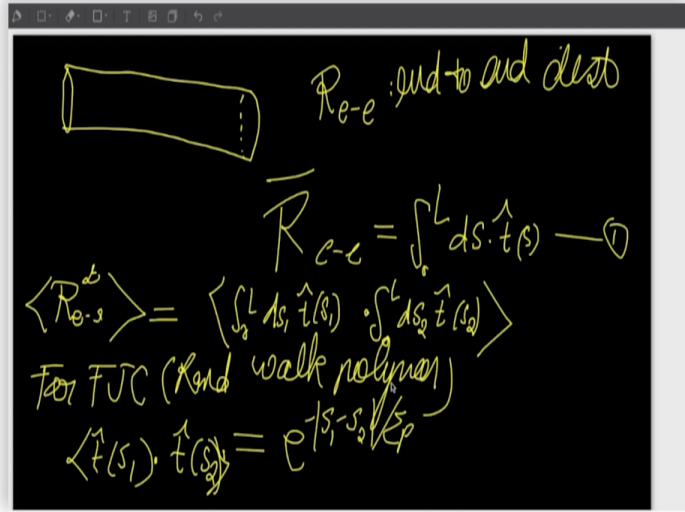
So, if we assume an actual rod sort of cylinder maybe bent at times, but then the end to end distance of rod can just simply be given as the integral over the whole length of every segment of that length times the tangent to that system.

$$\overline{R}_{e-e} = \langle \int ds \cdot t^\wedge(s) \rangle$$

So, when ds is small, so, this assumption is the ds are small segments. And the tangents are at every successive segment length of s, then the average squared end to end distance for such a rod can also be written as,

$$\langle R^2_{e-e} \rangle = \langle \int ds_1 t^\wedge(s_1) \int ds_2 t^\wedge(s_2) \rangle$$

 the integral over length L, ds_1 for some segment s1 times the unit vector of the tangent at that position integrated over the same thing for the segment second segment.

And this becomes relevant when our structure is not any more a straight rod because then this just simply holds true and is nothing more to it what when there is a random of polymer. So, for freely jointed chain or random or polymer $t^\wedge(s1)$ times $t^\wedge(s2)$ we said by definition is equal to,

$$\langle t^\wedge(s_1) \cdot t^\wedge(s_2) \rangle = e^{\frac{-|s_1 - s_2|}{\xi_p}}$$

e to the power minus magnitude of the difference between the two segments by Cp in negative exponential terms.

(Refer Slide Time: 28:29)

So, we want to try to do something with this expression above. And for that we split the integral into 2 conditions, one where s1 is greater than s2, and for obvious reasons where the opposite is true s1 is less than s2. Since we assume that these have equal contributions we can just simply rewrite the R_{e-e} squared averaged distance as

$$\langle R^2_{e-e} \rangle = 2 \int ds_1 \int ds_2 \frac{e^{-|s_1 - s_2|}}{\xi_p}$$

two times a common integral 0 to L ds_1, s1- s2 the difference in the limit that the control length is much greater than Cp which is reasonable we can subtract Δs from the upper and lower limits over here and end up with something that looks like this, which is

$$\langle R^2_{e-e} \rangle = 2 \int ds \int d\Delta s e^{-\Delta \frac{s}{\xi_p}}$$

But because and this is all because $l - \Delta s$ is approximately equal to L and true when L is much much greater than persistence length leaving us with an simplification of end to end distance squared average

$$\langle R^2_{e-e} \rangle = 2L\xi_p$$

But, we know that L is proportional to N that is from random walks statistics L is equal to, sorry L is proportional to N and R^2 the end to end distance is proportional to N because as I was trying

to write not L but R^2 is equal to Na^2. Therefore, R^2 average end to end is equal to La this is obvious this is the contour length is equal to from $2L\xi_p$ here therefore a is equal to $2\xi_p$.

So, this answers our question, what is the relationship between Kuhn's length and persistence length?

(Refer Slide Time: 32:09)

So, we wanted to define one more term so that was the radius of gyration R_G and we know that the R_G and this is also in the tutorial in the assignment the squared value of R_G is nothing but the average of the over all segments, all positions or if you have to prove what atoms than atoms.

$$\langle R_G^2 \rangle = \frac{1}{N} \Sigma \langle \left(\overline{R}_i - \overline{R_{cm}} \right)^2 \rangle$$

So what is R_{CM}? It is nothing but the position vector corresponding to the center of mass. By definition, R_{CM} the vector R_{CM} is equal to,

$$\overline{R_{cm}} = \frac{1}{N} \Sigma \overline{R}_i$$

in other words it is nothing but the average vector. So, you are subtracting you are finding the deviation square deviation of the average vector from every atom and averaging it, it turns out that you can write this the average radius of gyration or the root mean square radius of gyration for a random walk polymer as

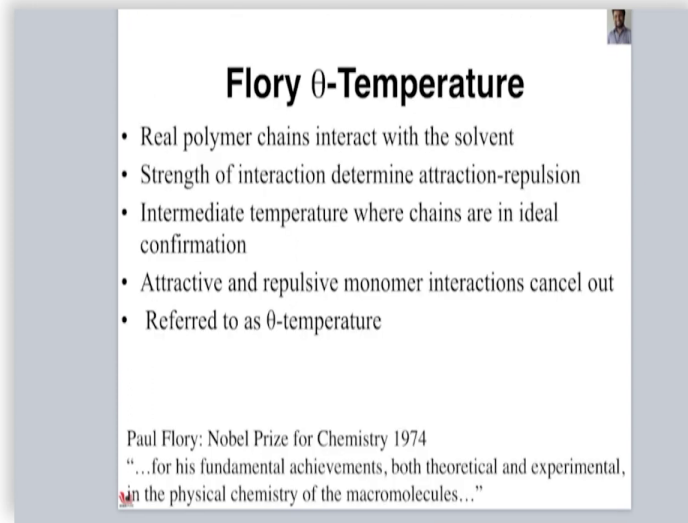$$\sqrt{\langle R_G^2 \rangle} = \sqrt{\frac{L\xi_p}{3}}$$

and I leave this for you to try to prove if you want to test it suffice to say we are going to use this number as a result. In order to find some order of magnitude estimates of radius of gyration we can indeed use the fact that experimentally measured values of Cp for DNA are 147 base pairs which is 150 base pairs approximately which is approximately 15 nanometers.
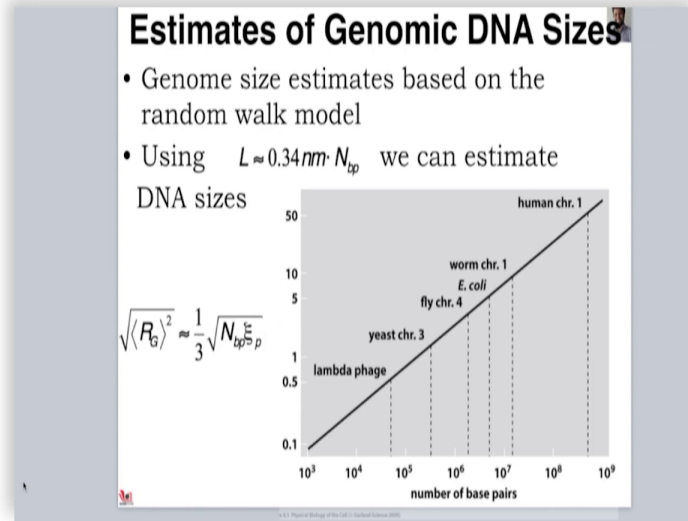
And for proteins, this is approximately a nanometer. So, we return to a concept from persistence length and radius of gyration to a Flory temperature before which I have now basically taken from the textbook a plot of the solution to the radius of gyration arithmetic by substituting L the length of the polymer as 0.34 nanometers times base pairs therefore $\frac{1}{3}\sqrt{N_{bp}\xi_p}$ for DNA alone and here you are looking at a log log plot of the radius of gyration on the y axis base pair lengths on the x axis. And on this log log plot these data sets seem to fall on a straight line with lambda phase at the lower limit and human chromosome one at the upper limit.

(Refer Slide Time: 36:09)

There is an issue with the fact that the end to end distance is related in real polymers to a scaling factor referred to as the Flory temperature because real polymer chains are known to interact with the solvent and the strength of interaction determines the attraction and repulsion. So, intermediate temperatures where the chains are in an ideal conformation where attractive and repulsive monomer interactions cancel out is then referred to as Flory temperature Paul Flory was again a Nobel laureate for his fundamental achievements in the physical chemistry of macromolecules in some sense is a pioneer of polymers.

(Refer Slide Time: 36:52)

Flexibility of Ideal Chains

- Tetrahedral angles of polymers such as polyethylene θ=68°
- Torsion angles determine flexibility
- n: number of bonds
- l: bond length
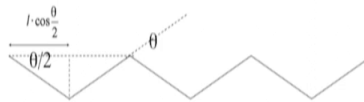
So, the flexibility of ideal chains can be considered with very simple non-biological chains such as polyethylene which have canonical 68 degrees between each chain position and the torsion angles determine flexibility number of bonds and bond lengths are then geometrically determined.

(Refer Slide Time: 37:18)

Largest End-End Distance

- Based on the skeleton of the polymer
- R_{max}: contour length of the chain

$$R_{max} = n \cdot l \cdot \cos\frac{\theta}{2}$$

And these can be then used to find the so called R_{max} value which is,

$$R_{max} = n \cdot l \cdot \cos\frac{\theta}{2}$$

geometrically n times the length of each rigid segment $\frac{cos\theta}{2}$ which is the angle between every segment.

(Refer Slide Time: 37:34)

And then the ideal bond vector relations gives us back for a freely jointed chain which has no correlation between bond vectors distant ones of $<R^2> = nL^2$ which is recovering our random walk of feely jointed chain statistics.

(Refer Slide Time: 37:51)

But for so called real chain bond vectors, there is a scaling factor of C_n which is Flory's characteristic ratio, indeed an infinite chain has n infinitely long obviously and then the mean square end to end distance of that becomes

$$\langle R^2 \rangle \approx C_\infty \cdot n \cdot l^2$$
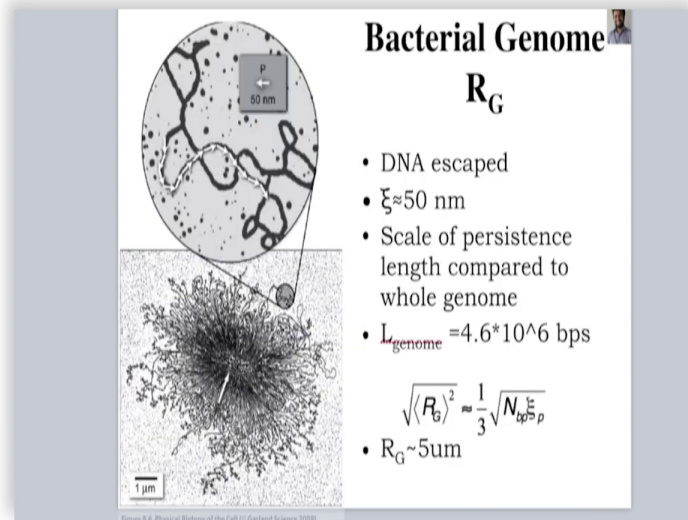
(Refer Slide Time: 38:17)

Bond-Lengths and Kuhn Lengths

- R_{max}: stretched length of the polymer
- $\langle R^2 \rangle$: Mean square end-to-end distance
- l: bond length
- a: Kuhn length
- C_∞: Flory's characteristic ratio

$$\langle R^2 \rangle = N \cdot a^2 = a \cdot R_{max} = C_\infty \cdot n \cdot l^2$$

Indeed the bond and Kuhn lengths are also then slightly modified by this Flory's characteristic ratio.

(Refer Slide Time: 38:26)

Bacterial Genome

$$R_G$$

- DNA escaped
- $\xi \approx 50$ nm
- Scale of persistence length compared to whole genome
- $L_{genome} = 4.6*10\text{^}6$ bps

$$\sqrt{\langle R_G \rangle^2} \approx \frac{1}{3}\sqrt{N_{bp}\xi_p}$$

- $R_G \sim 5$um

I said that you can get at a order of magnitude of persistence length with DNA but you can also get at a radius of gyration for DNA and remember we said that random of polymers of freely jointed chains always want to come to the center. In other words they are highly dense at the center the probability of their density is highest in center it decays in a Gaussian fashion in 1D, 2D and 3D. So, it is interesting to note that when you when in the image down below in an electron micrograph the cell was lysed and the DNA was stained in electro micrograph then the with a persistence length of 50 nanometers the radius of gyration comes to 5 microns and remember we always talk about the scale of an *E. coli* cell this is about 2 microns it is the longest size of a non-dividing cell in our dividing cell you can claim 4 microns.

In effect, circularly the DNA that comes out of a burst *E. coli* cell is much larger than the cell itself, what this should tell you that not just an eukaryotes but even in prokaryotes somehow the DNA is packed, packed and organized so that it can nicely fit in and perform its functions despite being packed so tightly.
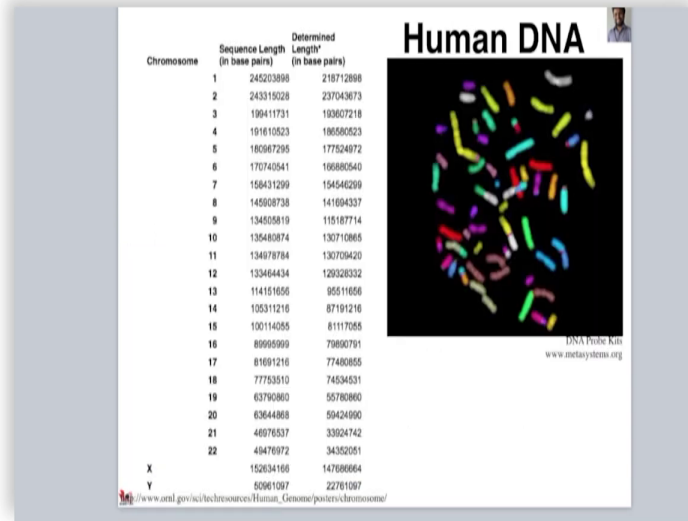
And this is now a simple list of sequence lengths and of human DNA just to remind you that these vary quite a lot and individual chromosomes, yes individual chromosomes of the human genome at least have 3, 6, 10 to the power 9 by spares order of magnitude lengths 0.1, 0.2 or even higher in some cases.
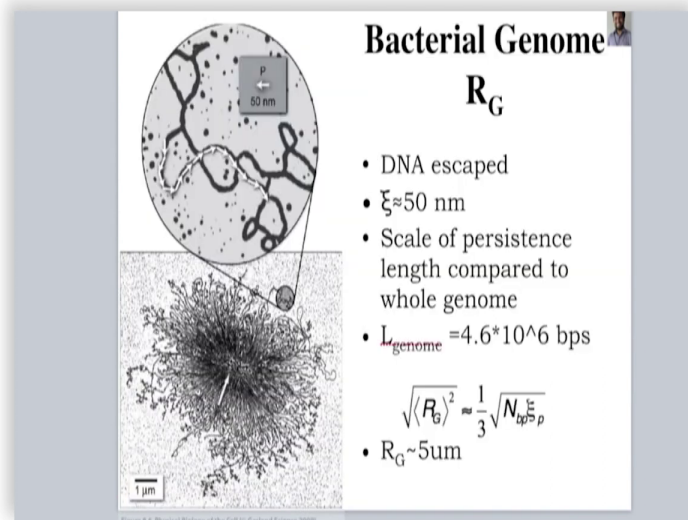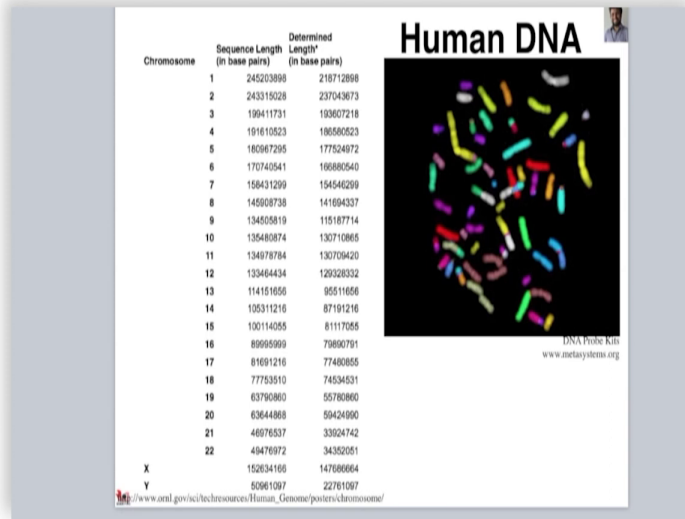
And this is quite fascinating because if I told you that the genome length of bacteria has a radius of gyration of 5 microns as a free polymer and you should be able to see this on the scale bar also

by the way this is not just speculation or theory this is coming all sort of experiments in other words reconciliation experiments in theory.

(Refer Slide Time: 40:43)

So in the same way the human genome is going to be much longer you are free to calculate this and I am going to ask you this question on Thursday, what is the average R_G of the human genome? You will have to make some assumptions but I would like you to do this arithmetic anyway.

(Refer Slide Time: 41:05)
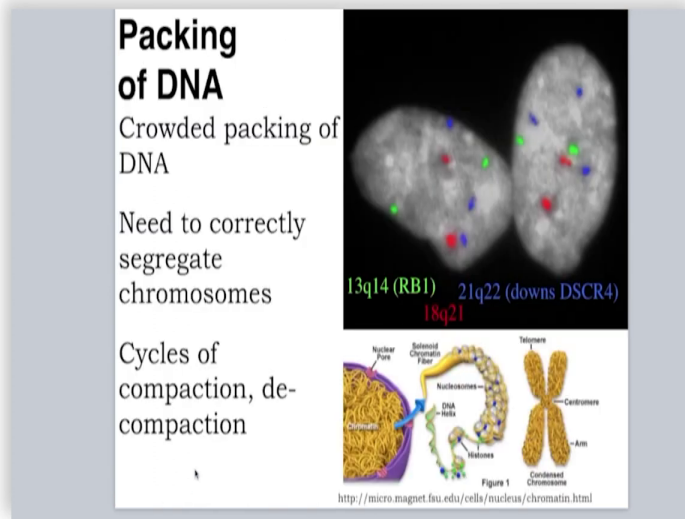
As it turns out DNA in eukaryotes and especially in humans which has been very well studied is tightly packed and you are tightly packing it but when you segregate it, it kind of decondenses or rather it condenses even further and it has to be correctly segregated and the cycles of compaction decompaction you have to do it right, I mean you need to get it in the correct place because as you see in the image over here the blue marker indicates what happens when you do not correctly segregate it and you end up with more than one copy of a chromosome this indeed is an is an unfortunate case of a disease called down syndrome.

(Refer Slide Time: 41:50)
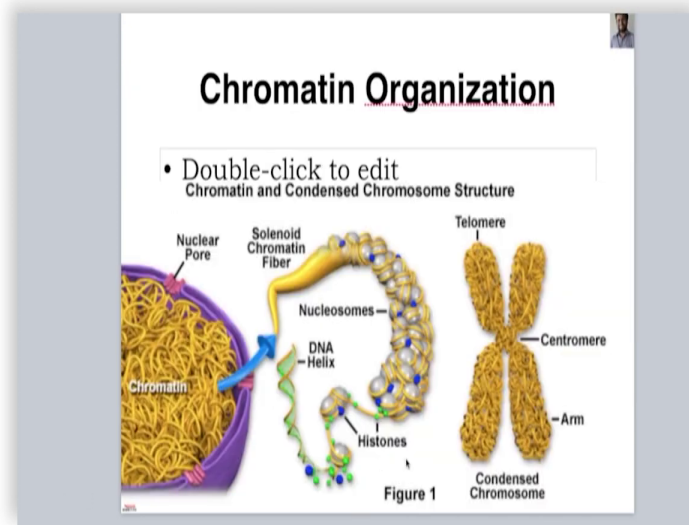
So, chromatin organization is indeed seen to involve the packing of I mean this noodles Chinese noodle style appearance of DNA into increasingly more complex structures so when we return we are going to discuss, what these complex structures are? What are their dimensions? What consequences do they have for our understanding of human genome and its organization? And how far are we with a complete understanding of packing with a connection back to polymer theory. So, thank you very much for listening and I am going to stop.