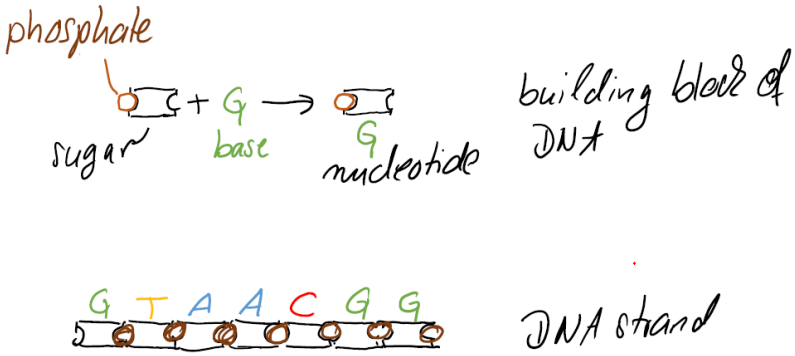
1. The universal features of cells on earth

It is estimated that there are more than 10 million living species on Earth today. Most living organisms are single cells. Others, such as ourselves, consist of groups of cells performing specialized functions. The whole organism has been generated by cell divisions from single cell. The single cell, therefore, is the vehicle for all of the hereditary information that defines each species.

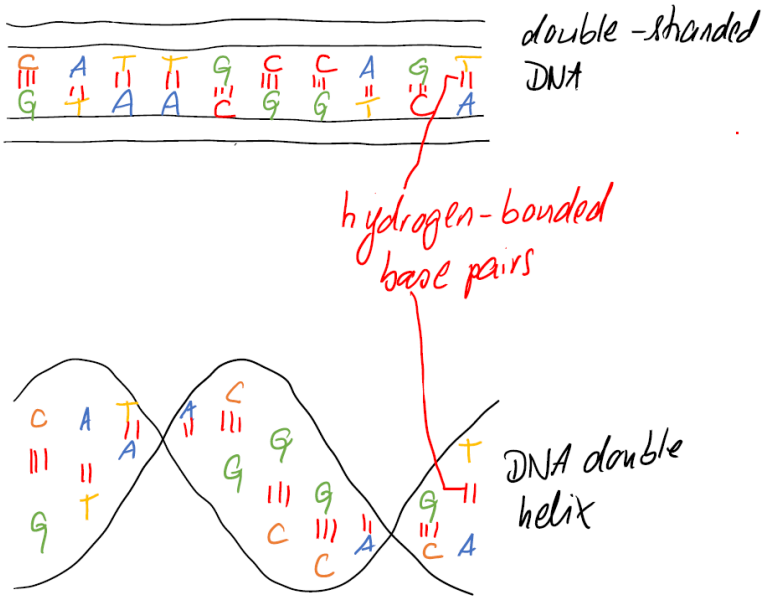
This chapter summarizes the introductory part of the book Molecular Biology of the Cell by Bruce Alberts et al [1].

1.1. All cells store hereditary information in DNA

Cells, like computers, store information. All living cells store their hereditary information in the form of double-stranded molecules of DNA, polymer chains formed always of the same four types of monomers. These monomers, chemical compounds known as nucleotides, are strung together in a linear sequence that encodes the genetic information. The nucleotides consist of a sugar-phosphate molecule with a base attached to it. The bases are of four types: A (adenine), G (guanine), C (cytosine), T (thymine). A single strand of DNA consists of nucleotides joined together by sugar-phosphate linkages. We can take a piece of DNA from human cell and insert it into a bacterium and the information will be successfully read, interpreted, and copied.

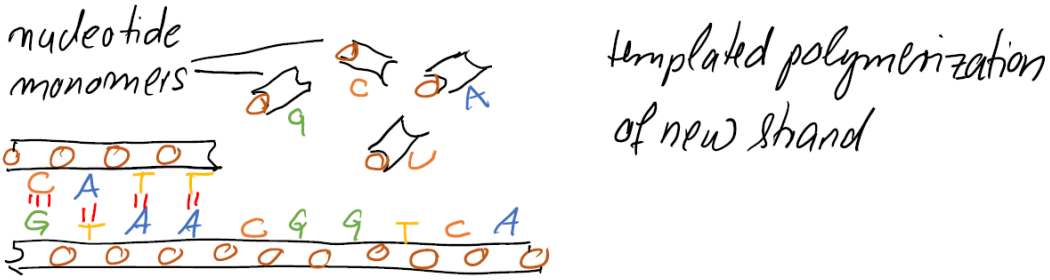


A normal DNA molecule consists of two complementary strands. The nucleotides within each strand are linked by strong (covalent) chemical bonds; the complementary nucleotides on opposite strands are held together more weakly, by hydrogen bonds. We will consider these different types of bonds in chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**. The two strands twist around each other to form a double helix.

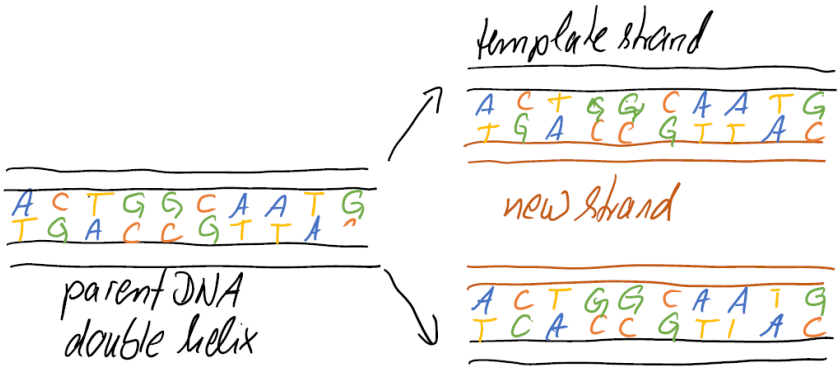


1.2. All cells replicate their hereditary information by templated polymerization

DNA is synthesized on a template formed by a preexisting DNA strand. In other words, through templated polymerization, the sequence of nucleotides in an existing DNA strand controls the sequence in which nucleotides are joined together in a new DNA strand; T in one strain pairs with A in the other, and G in one strand with C in the other.

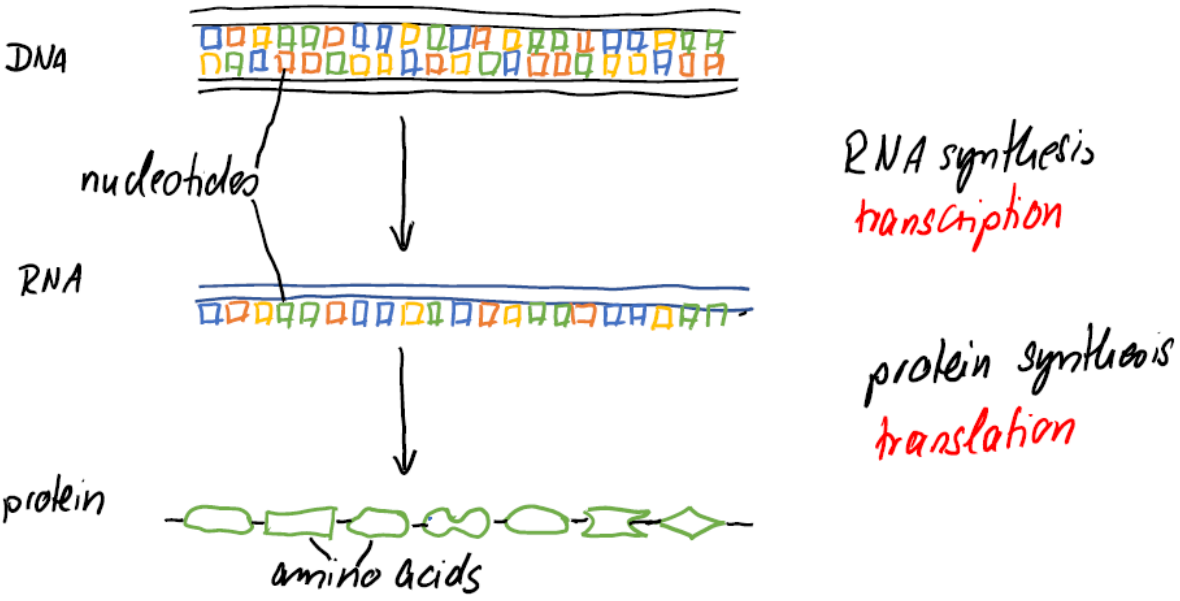


The bonds between the base pairs are weak compared with the sugar-phosphate links, and this allows the two DNA strands to be pulled apart without breakage of their backbones. Each strand can then serve as a template for the synthesis of a fresh DNA strand. This process is called DNA replication and is universal for all organisms.

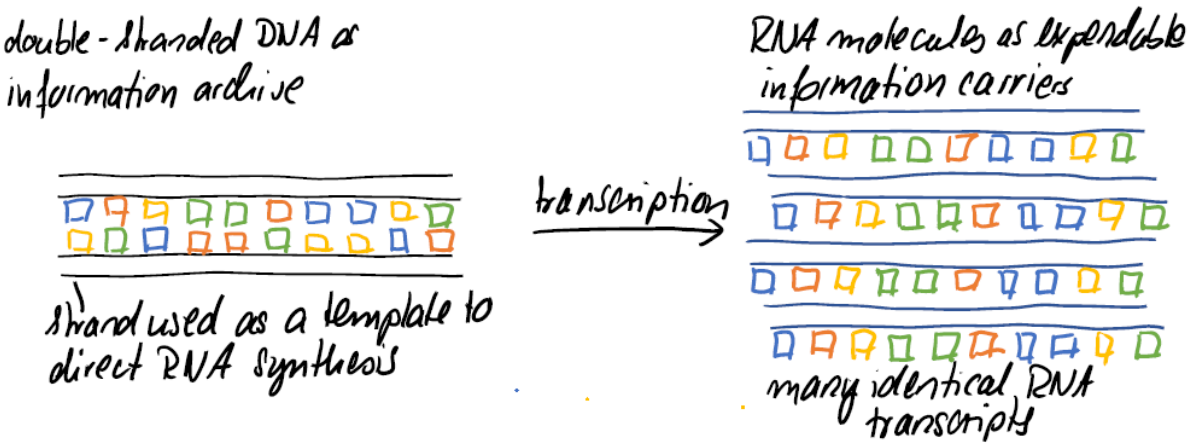


1.3. Gene expression: All cells transcribe and translate portions of their hereditary information

DNA must express its information, leading to the production of two other key classes of polymers: RNA and proteins. Gene expression begins with a templated polymerization called transcription, in which segments of the DNA sequence are used as templates for the synthesis of shorter molecules of RNA. Later, in the process of translation, these RNA molecules direct the synthesis of polymers called proteins.



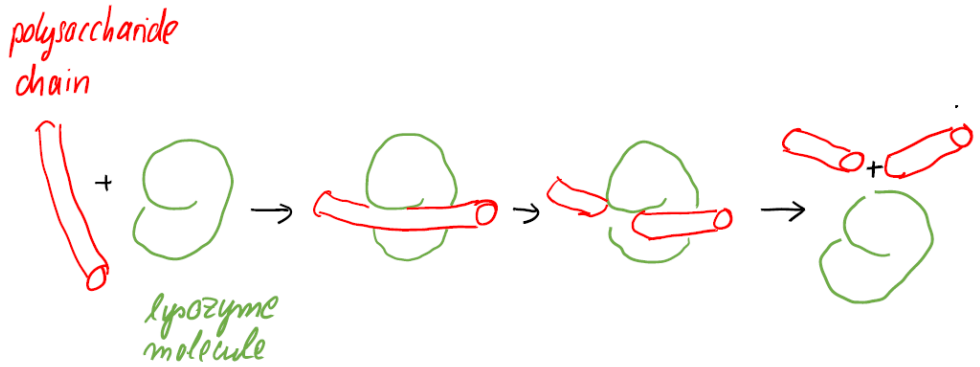
Each cell contains a fixed set of DNA molecules - its archive of genetic information. A given segment of this DNA guides the synthesis of many identical RNA transcripts, which serve as working copies of the information stored in the archive. Many different sets of RNA molecules can be made by transcribing different parts of a cell's DNA sequences, allowing different types of cells to use the same information store differently.



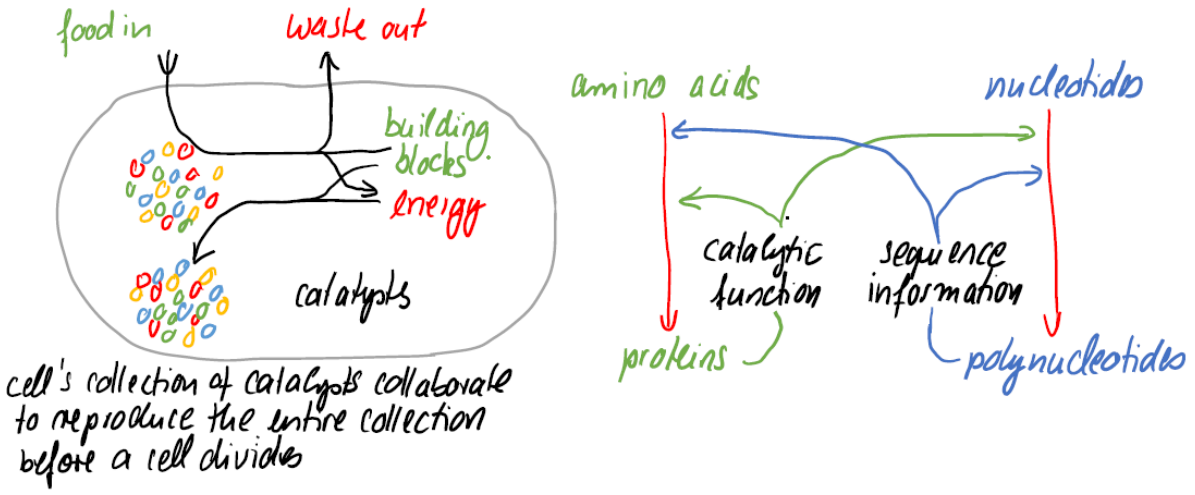
Like DNA and RNA, proteins carry information in the form of a linear sequence of symbols. There are 20 types of monomers of protein called amino acids. The information in the sequence of a messenger RNA molecule is read out in groups of three nucleotides at a time: each triplet of nucleotides, or codon, specifies a single amino acid in a corresponding protein. Since the number of distinct triplets that can be formed from four nucleotides is 4^3, there are 64 possible codons, all of which occur in nature. However, there are only 20 occurring in natural amino acids. Several codons correspond to the same amino acid.

1.4. All cells use proteins as catalysts

Each of the protein molecules is a polypeptide, created by joining its amino acids in a particular sequence. Through evolution, this sequence has been selected to give the protein a useful function. Thus, by folding into a precise three-dimensional form with reactive sites, these polymers can bind with high specificity to other molecules and can act as enzymes to catalyze reactions that make or break chemical bonds. In this way they direct the vast majority of chemical processes in the cell. In chapter **Fehler! Verweisquelle konnte nicht gefunden werden.** we will discuss the interactions that control the three-dimensional structure of proteins.



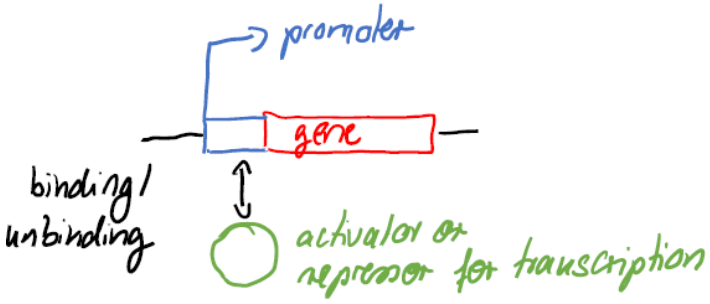
Protein have many other functions including maintaining structures, generating movements, or sensing signals.



From the most fundamental point of view, we can consider the cell as a self-replicating collection of catalysis. Polynucleotides (the nucleic acids DNA and RNA, which are nucleotide polymers) provide the sequence information, while proteins (amino acid polymers) provide most of the catalytic functions that serve (through a complex set of chemical reactions) to bring about the synthesis of more polynucleotides and proteins of the same types.

1.5. What is a gene?

DNA molecules contain specifications for thousands of proteins. Special sequences in the DNA define where the information for each protein begins and ends. Each such DNA segment represents one gene.



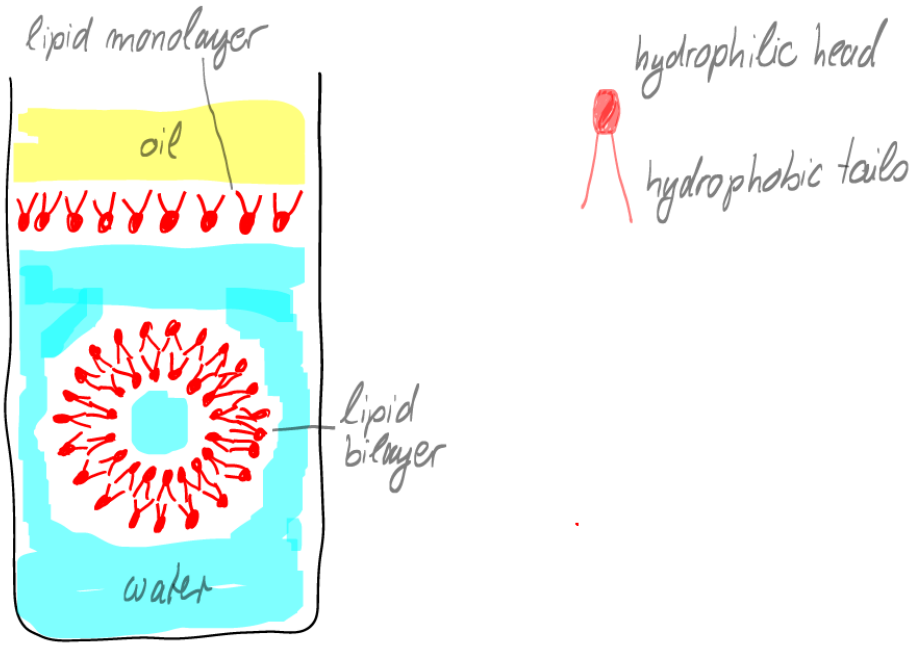
In all cells, expression of individual genes is regulated: the cell adjusts the rate of transcription and translation of different genes independently, according to need. Stretches of regulatory DNA are interspersed among segments that code for a protein, and these noncoding regions bind to special protein molecules that control the rate of transcription. In this way, the genome of the cell - that is, the totality of its genetic information as embodied in its complete DNA sequence - dictates not only the nature of the cell's proteins, but also when and where they are to be made.

1.6. Life requires free energy

A living cell is a dynamic chemical system, operating far from chemical equilibrium. For a cell to grow or to make new cells, it must take in free energy from the environment, as well as raw materials, to drive the necessary synthetic reactions. This consumption of free energy is fundamental to life. When it stops, a cell decays toward chemical equilibrium and soon dies.

1.7. All cells are enclosed by membrane

Each cell is enclosed by a membrane - the plasma membrane. This container acts as a selective barrier that enables the cell to concentrate nutrients and retain the products it synthesizes, while excreting its waste products.



The molecules that form a membrane are amphiphilic - that is, they consist of one part that is hydrophobic (water-insoluble) and another part that is hydrophilic (water-soluble). In particular, phospholipids have a hydrophilic head and a hydrophobic tail. At an interface between oil and water, they spontaneously form a single sheet with their head groups facing the water. When immersed in water, they self-assemble to form bilayers enclosing aqueous compartments.

If a cell is to grow and reproduce, it must be able to import raw materials and export waste. All cells therefore have specialized proteins embedded in the membrane that transport specific proteins from one side to the other.

1.8. A living cell can exist with fewer than 500 genes

What are minimum requirements for a living cell? The bacterium *Mycoplasma genitalium* has one of the smallest known genomes. It lives as a parasite in mammals, and its environment provides it with many of its molecules. It has about 530 genes, about 400 of which are essential.

Its genome of 580,070 nucleotide pairs represents 145,018 bytes of information. The minimum number of genes for a viable cell in today's environment is probably not less than 300, with only about 60 genes in the core set that is shared by all living species.

1.9. Molecular biology began with a spotlight on *Escherichia coli*

Many cell types can be cultivated and analyzed under laboratory conditions. The "working horse" of molecular biology is the bacterium *Escherichia coli*. It lives in the gut of every human, is easy to cultivate in the laboratory with a generation time of ~ 30 min. Because genetic manipulation is comparatively simple, work with *E. coli* has led to the discovery of many basic principles in molecular biology. Many of these principles are shared by animal cells although they tend to be more complex.

Length	2 µm
Volume	1 μm^3 = 1fL
Surface area	6 μm^2
Weight	1 pg
Dry weight	30 %
Number of proteins per cell	$2.4 \cdot 10^6$
Number of DNA molecules per cell	2

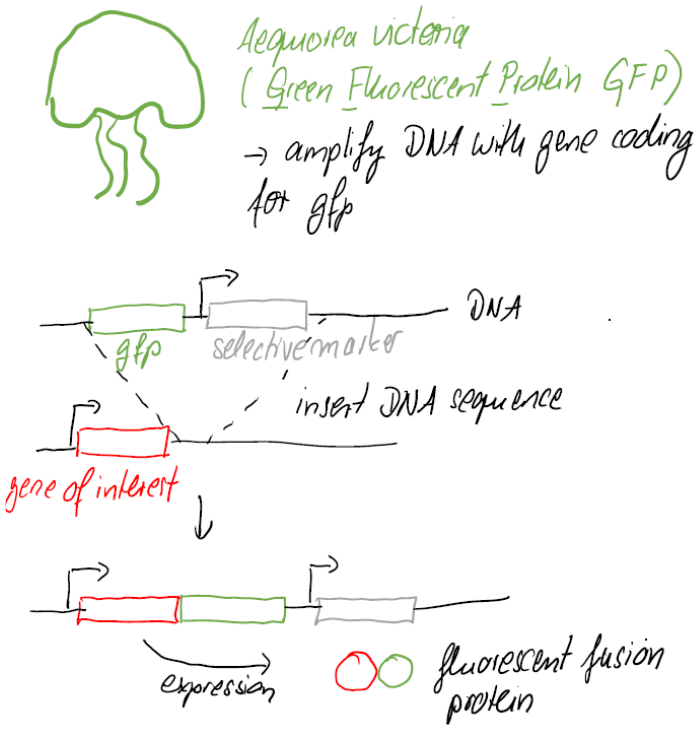
The bacterial standard ruler. Estimated values for *E. coli*.

It is important to be aware that the cell is a crowded environment. Assuming that the radius of a typical protein is r ≈ 2 nm, please estimate the volume fraction of an *E. coli* cell taken up by proteins only. Crowdedness affects the rate of chemical reactions and the diffusion constant discussed later in this course.

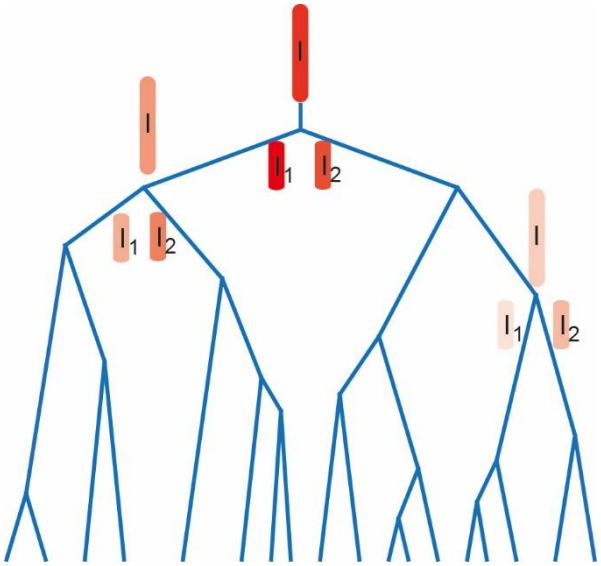
1.10. Experiment: Counting molecules in a living cell

Different proteins are maintained at very different concentrations. For example, different genes are expressed (see 1.3) at very different rates. As a consequence, the number of messenger RNA molecules as well as the number of proteins that exist within a cell varies by orders of magnitude. While a number of biochemical methods exist that allow determining relative concentrations of proteins within cells, it is tricky to determine the absolute number (or concentration) of a specific type of protein per cell. One method relies on the fact, that during cell division, molecules partition randomly into both daughter cells. Here, we will introduce an experiment that allows counting proteins in living cells [2, 3].

First step: Construction of a strain expressing a fusion protein between the protein of interest and a fluorescent reporter.

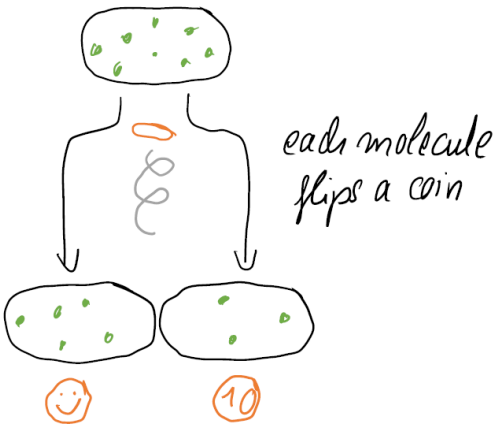


Second step: Perform single-cell fluorescence microscopy and analyse the fluorescence intensity of mother and daughter cells.



Third step: Use binomial distribution to determine the number of fluorescent molecules.

As cells grow in length, the total number of proteins increases due to continuous gene expression. Once a bacterium has grown a predefined length, it divides and forms two daughter cells of approximately same length. During cell division, the content of the mother cell must be partitioned into the daughter cells. While equal partitioning of chromosomal DNA and some low-copy plasmids is controlled by the cell, partitioning of mRNA and proteins usually occurs by random partitioning.



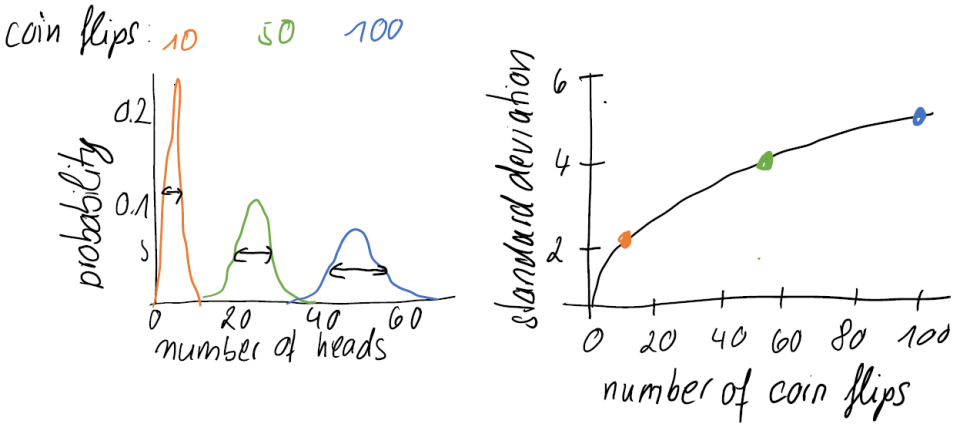
Importantly, the fluctuations in protein number that arise due to partitioning of mother cells can be used to quantify the total number of a proteins species. Conceptually, the passive process of partitioning can be thought of as a series of coin flips [4]. We assume that the number total of proteins N is constant. The probability of going to daughter 1 is given by p and the probability of going to daughter 2 by $q = 1 - p$. Both outcomes are equally likely, i.e. $p = q = ½$. To be precise, we are facing a series of equal and independent experiments that can have only two results. The probability $P(n_1, N)$ of having n_1 of our N molecules in daughter 1 is given by the binomial probability distribution

$$P(n_1, N) = \binom{N}{n_1} p^{n_1} q^{N-n_1}$$

We can calculate the expectation value and the variance to be

$$Exp(n_1) = \langle n_1 \rangle = \sum_{n_1=0}^{N} n_1 \binom{N}{n_1} p^{n_1} q^{N-n_1} = \cdots = Np$$

$$Var(n_1) = \langle {n_1}^2 \rangle - \langle n_1 \rangle^2 = \sum_{n_1=0}^{N} {n_1}^2 \binom{N}{n_1} p^{n_1} q^{N-n_1} - (Np)^2 = \cdots = Nqp$$



By experimentally determining the number difference between the daughter cells $\langle(n_1 - n_2)^2\rangle$, the total number of molecules N can be determined.

$$\langle(n_1 - n_2)^2\rangle = \langle\left(n_1 - (N - n_1)\right)^2\rangle = \cdots = N$$

If the number of molecules per cell can be counted by single molecule analysis, then this relation can be verified. Using reporter constructs for mRNA in *E. coli*, binomial partitioning has indeed been proven [5]. Usually single molecules cannot be resolved. However, by determining the difference in fluorescence intensities of cells generating green fluorescent proteins (GFP), the total number of GFP molecules can be counted after a calibration step [2].

The microscopic images of the dividing cells contain fluorescence intensities for each cell in arbitrary units. The measured intensity of each cell (i.e. the number of detected photons per cell) strongly depends on the specifics of the experimental setup. Many factors, such as optics of the microscope, CCD chip (camera) and excitation source strongly alter the number of collected photons. Therefore, it is not possible to find a 'global' calibration to calculate how many proteins are involved in generating the signal. The calibration factor α which connects the observed intensity to the number of fluorescent proteins is defined by

$$I_{tot} = \alpha N_{tot}$$

The knowledge of the binomial process and the definition of α allow us to calculate that

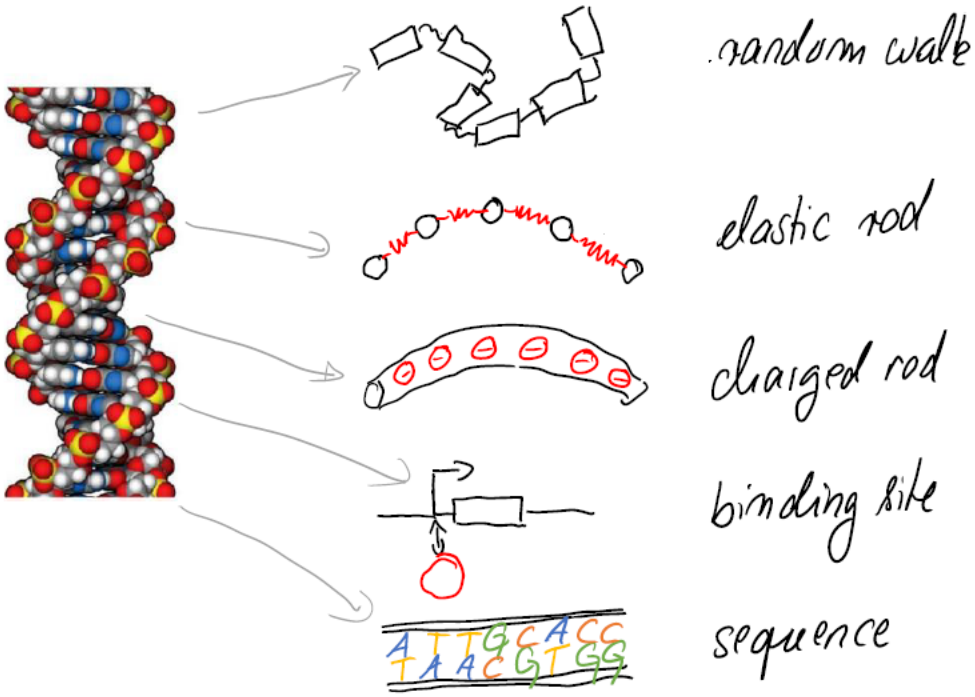
$$\langle(I_1 - I_2)^2\rangle = \cdots = \alpha I_{tot} = \alpha(I_1 + I_2)$$

assuming that $I_1 + I_2 = I_{tot}$ and $N_1 + N_2 = N_{tot}$. By plotting the absolute value of $I_1 - I_2$ as a function of I_{tot} we can easily determine the calibration factor and calculate the number of proteins.

2. DNA: from equilibrium conformation to virus packaging

Given different biophysical aspects addressed in the course, we will view molecules and cells from many different angles. Here, we consider deoxyribonucleic acid (DNA). Although we know the structure of DNA at atomic resolution, coarse-grained descriptions of DNA will be used for the topics considered in this lecture. In this chapter, we will focus on conformations and elastic properties of DNA molecules. We will start by describing DNA as a polymer performing a random walk. The simple random walk model captures many important features of DNA including its equilibrium size. However, when mechanical forces are applied to the molecule, deforming it, we will find out that bending energies need to be considered for describing elastic properties of the DNA molecule. Such mechanical forces are at play, for example, when DNA-binding proteins form loops, when DNA is packaged into viruses, or when DNA is compacted to fit into the cellular nucleus of eukaryotic cells. In chapter 2.3, DNA will be considered as a physical entity, more specifically as a polyelectrolyte. Gene regulation will

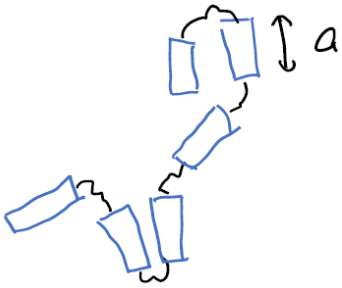
be addressed in chapters XX and in Biological Physics 2. There, often DNA is depicted as a sequence of blocks where each block has specific function, e.g. a gene or a regulatory sequence. For molecular cloning or when addressing evolutionary questions, the sequence of DNA is of importance.



2.1. The Freely-Jointed-Chain or Random Walk model

2.1.1. The end-to-end distance

When visualized using a microscope, DNA assumes the structure of a random coil. Here, we aim at describing the mean end-to-end distance and the free energy required for deforming the polymer. To this end, we discretize the polymer into a series of segments (Kuhn segments) with length *a*.



For simplicity, the model is confined to one dimension. The probability to step to the right p_r equals the probability to step to the left p_l

$$p_r = p_l = 1/2$$

All steps are independent, there is no memory. For a chain with N segments there are 2^N possible configurations each with a probability $1/2^N$.

After *N* steps, the expected value of the walker's distance *R* from the origin is

$$\langle R \rangle = \left\langle \sum_{i=1}^{N} x_i \right\rangle \text{ with } x_i = \pm a$$

Since $\langle x_i \rangle = 0$, the average displacement of the random walker after *N* steps is $\langle R \rangle = 0$. A more useful measure of the walker's departure from origin is to examine the mean squared displacement. The variance of the probability distribution of R is

$$\langle R^2 \rangle = \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \, x_j \right\rangle$$

The significance of the standard deviation $\sqrt{\langle R^2 \rangle}$ is as follows: the probability of finding our random walker within one standard deviation is close to 70%*;* thus, it is a good surrogate of the typical size of a polymer.

$$\langle R^2 \rangle = \sum_{i=1}^{N} \langle x_i^2 \rangle + \sum_{i \neq j=1}^{N} \langle x_i x_j \rangle = Na^2$$

The end-to-end distance is therefore

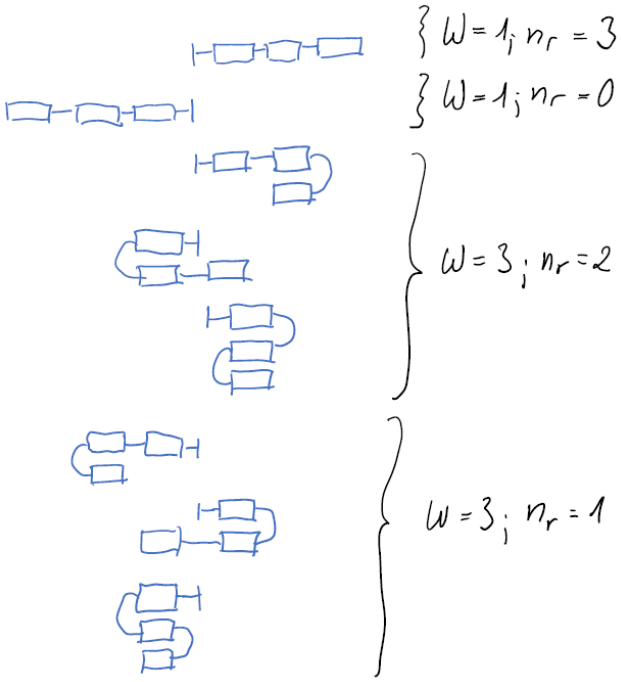
$$\sqrt{\langle R^2 \rangle} = \sqrt{N} a$$

2.1.2. The end-to-end probability distribution

Next, we will describe the end-to-end probability distribution for a one-dimensional polymer. The probability of a given configuration depends upon its microscopic degeneracy. In the case in which the walker makes a total of N steps: What is the probability that n_r of those steps will be to the right and hence $n_l = N\text{-}n_r$ to the left?

There are W distinct ways of achieving an end-to-end distance with n_r steps to the right

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



The probability of an overall departure n_r from the origin is

$$p(n_r; N) = \frac{N!}{n_r!\,(N - n_r)!}\left(\frac{1}{2}\right)^N$$

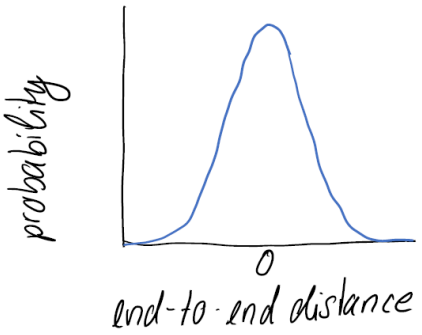
You may confirm normalization through the binomial theorem. Next, we calculate the probability distribution function of the end-to-end distance $\mathrm{R} = (\mathrm{n}_r - \mathrm{n}_l)\mathrm{a}$ with $\mathrm{n}_r + \mathrm{n}_l = N$

$$p(R; N) = \frac{N!}{\left(\frac{N}{2} + \frac{R}{2a}\right)!\left(\frac{N}{2} - \frac{R}{2a}\right)!}\left(\frac{1}{2}\right)^N$$

Using Stirling's approximation and Taylor expansion for large N, the probability distribution function of the end-to-end distance (1D) is

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

for $R \ll Na$.



This result is a special case of the central limit theorem: The probability distribution of $x_1 + x_2 + \ldots + x_N$ (sum of identically distributed independent variables) is Gaussian in the limit of large

N, as long as the mean and variance of each individual x_i are finite. It follows that $\langle \vec{R} \rangle = 0$ and $\sqrt{\langle \vec{R}^2 \rangle} = \sqrt{N}a$ regardless of dimension.

The probability distribution function of the end-to-end distance (3D) is

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

$P(R;N)$ is sharply peaked around R = 0. It follows that entropy determines the elastic properties of polymer chains such as DNA. When stretching out a polymer, it quickly relaxes to the state with R ≈ 0.

Estimate: End-to-end probability for the *E. coli* genome

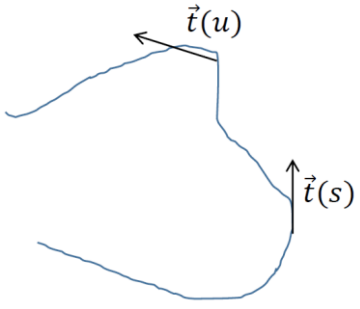
Genomic DNA of *E. coli* is roughly 5 Mbp long. The Kuhn segment length of DNA is a_{DNA} = 300 bp. Thus N = 15000. Estimate the probability for R = 0, R = 500a, R = 1000a.

2.1.3. The persistence length

With the random walk model, we can describe the structure of long polymers, whose contour length *L* is much longer than the persistence length ξ, which is the length over which the polymer is essentially straight. In particular, the persistence length is the length scale over which the tangent-tangent correlation function decays along the chain.

$$\langle \vec{t}(s)\vec{t}(u) \rangle = e^{-\frac{|s-u|}{\xi}}$$

where $\vec{t}(s)$ is the tangent vector evaluated at a distance s along the polymer and the notation $\langle ... \rangle$ is an instruction to average over all the configurations.



How is the persistence length related to the Kuhn segment length of the FJC model? We calculate the mean-squared end-to-end distance of an elastic beam undergoing thermal fluctuations, and compare it with the same quantity obtained for the FJC. The end-to-end vector $\vec{R}$ of a beam can be expressed in terms of the tangent vector $\vec{t}(s)$

$$\vec{R} = \int_0^L ds\, \vec{t}(s)$$

As a result, we can write

$$\langle \vec{R}^2 \rangle = \left\langle \int_0^L ds\, \vec{t}(s) \int_0^L du\, \vec{t}(u) \right\rangle = 2\int_0^L ds \int_s^L du\, e^{-\frac{|u-s|}{\xi}}$$

We change variables $x = u - s$ and replace the upper bound of integration $L - s$ by ∞.

In the limit $L >> x$

$$\langle \vec{R}^2 \rangle \approx 2\int_0^L ds \int_0^\infty dx\, e^{-\frac{x}{\xi}} = 2L\xi$$

Thus

$$a = 2\xi$$

How big is a genome? The radius of gyration is a useful measure of polymer size. It measures the average distance between monomers *i* and the center of mass *cm* of the polymer

$$\langle {R_G}^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \left\langle \left(\vec{R}_i - \vec{R}_{cm} \right)^2 \right\rangle$$

with $\vec{R}_{cm} = \frac{1}{N} \sum_{i=1}^{N} \vec{R}_i$. One can show that

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

Importantly, the radius of gyration of a polymer scales like

$$\sqrt{\langle R_G^2 \rangle} \propto \sqrt{N}$$

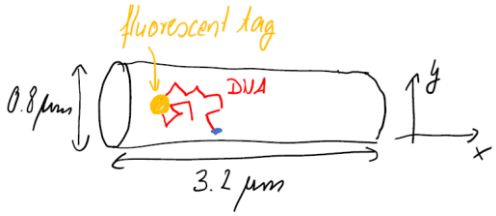
This result does not take into account excluded volume effects which are important for two-dimensional systems or for large *N*.

2.1.4. Effects of polymer confinement and tethering

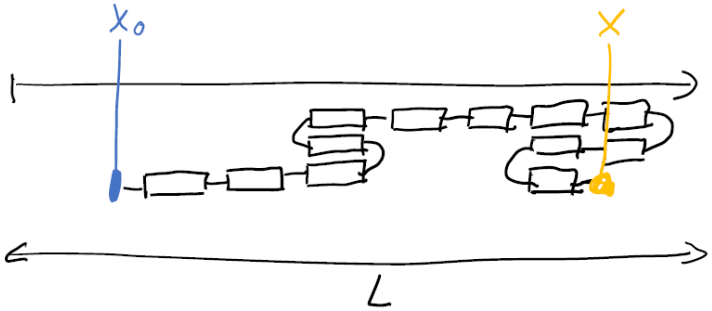
Estimate: The length of one base pair of DNA is 0.34 nm. The larger chromosome of *Vibrio cholera* has a length of 3 Mbp. Estimate the contour length of the DNA and the radius of gyration and compare to the size of a typical bacterium.

In living cells, DNA is confined to the cellular container. Moreover, there is experimental evidence that the chromosomes are tethered to the cell envelope. Here, we will assess the effects of confinement and tethering on the probability distribution of end-to-end distance. In the experiment motivating our model, the larger chromosome of the rod-shaped bacterium *Vibrio*

cholerae carried a fluorescent tag at a specific location of its chromosome. Using fluorescence microscopy, the distribution of positions of the tag was measured along the x-, and y-direction.



The chromosome position distribution was in agreement with a Gaussian in the x- but not in the y-direction, suggesting that the FJC model does not describe the distribution properly. To account for confinement and tethering, we again use a 1D random walk model, but keep only those walks which stay within the cell.



The fraction of random walks that start at $x = x_0$ and end at x without leaving the cell G satisfies the diffusion equation [4]. We will derive the diffusion equation later in chaper XXX.

$$\frac{\partial G(x;N)}{\partial N} = \frac{a^2}{2}\frac{\partial^2 G(x;N)}{\partial x^2}$$

The probability that a walk which stays in the cell also ends up at position x is

$$P(x;N) = \frac{G(x,N)}{\int_0^L G(x;N)dx}$$

Initial conditions:

$$G(0;N) = G(L;N) = 0$$

$$G(x,0) = \delta(x - x_0)$$

To solve the diffusion equation, we expand $G(x;N)$ into a Fourier series whereby each term satisfies the absorbing boundary condition

$$G(x;N) = \sum_{n=1}^{\infty} A_n(N) sin\left(\frac{n\pi}{L}x\right)$$

The initial condition states

$$\delta(x - x_0) = \sum_{n=1}^{\infty} A_n(0) sin\left(\frac{n\pi}{L} x\right)$$

and needs to be solved for the constants $A_n(0)$. To do this, we multiply both sides by $sin\left(\frac{m\pi}{L} x\right)$ and integrate from *0* to *L* yielding

$$A_m(0) = \frac{2}{L} sin\left(\frac{m\pi}{L} x_0\right)$$

We now substitute the Fourier expansion into the differential equation

$$\sum_{n=1}^{\infty} \frac{dA_n(N)}{dN} sin\left(\frac{n\pi}{L} x\right) = -\frac{a^2}{2} \sum_{n=1}^{\infty} A_n(N) \left(\frac{n\pi}{L}\right)^2 sin\left(\frac{n\pi}{L} x\right)$$

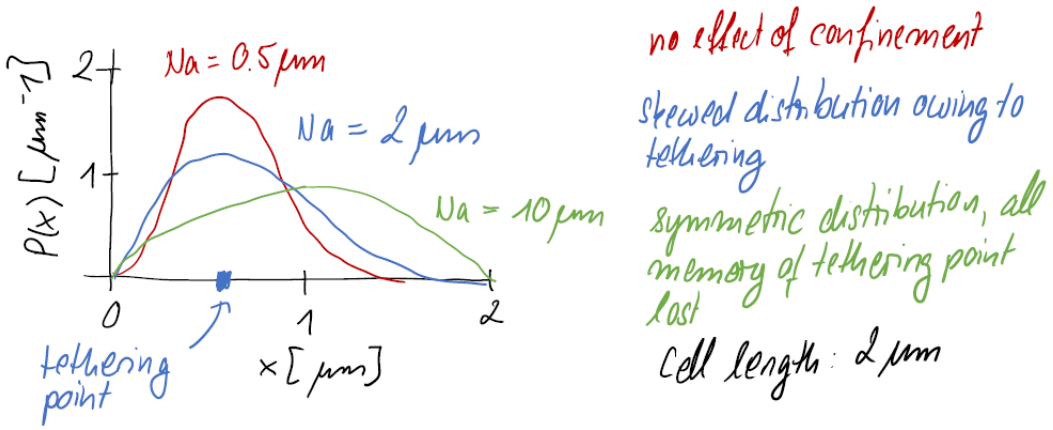
and use the same trick again for determining $A_m(N)$. Fraction of random walks that start at $x = x_0$ and end at x without leaving the cell G is

$$G(x; N) = \sum_{n=1}^{\infty} \frac{2}{L} sin\left(\frac{n\pi}{L} x_0\right) sin\left(\frac{n\pi}{L} x\right) exp\left(-\frac{a^2}{2}\left(\frac{n\pi}{L}\right)^2 N\right)$$

and the probability distribution of the end-to-end distance of a polymer confined in a cell

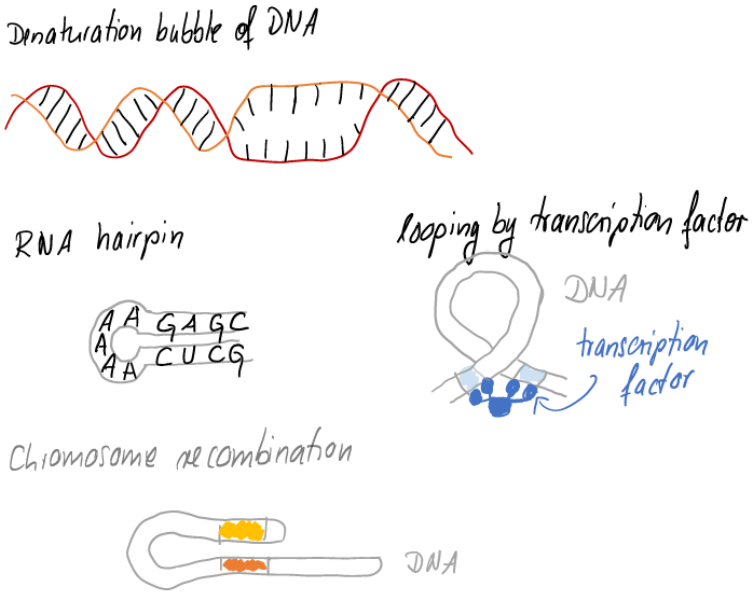
$$P(x; N) = \frac{1}{L} \frac{\sum_{n=1}^{\infty} \frac{2}{L} sin\left(\frac{n\pi}{L} x_0\right) sin\left(\frac{n\pi}{L} x\right) exp\left(-\frac{a^2}{2}\left(\frac{n\pi}{L}\right)^2 N\right)}{\sum_{n=1}^{\infty} \frac{2}{L} sin\left(\frac{n\pi}{L} x_0\right) \frac{1}{n\pi}(1 - cos(n\pi)) exp\left(-\frac{a^2}{2}\left(\frac{n\pi}{L}\right)^2 N\right)}$$

The distribution for different contour lengths between the tethering point and the fluorescent marker shows that confinement has little effect for short chains, intermediate lengths show a skewed distribution due to tethering at x = 0.75 µm. For very long chains, the distribution is again symmetric (but not Gaussian) with all memory of the tethering point lost.



2.1.5. DNA looping

Nucleic acids form loops in a variety of different settings. For example, double-stranded DNA can melt locally and form denaturation bubbles. This happens at high temperature, high pH, or when the RNA polymerase responsible for transcription 'reads' the information stored in DNA. RNA is usually single-stranded, but it forms complex three-dimensional structures because partially complementary regions form hydrogen bonds and thus create hairpin-like structures. As part of the process of gene regulation, some transcription factors form DNA-loops. Genetic recombination between distant parts of the chromosome are important for example in B cells (immune cells). The first step to recombination is, again, loop-formation.



Here, we will derive the entropic cost of loop formation which can be captured by the FJC model. In the next chapter, this model will be refined considering bending energies. Looping of large DNA fragments is dictated by the difficulty of distant ends finding each other. Looped DNA has fewer conformations available and thus looping confers an entropic cost. To see this, we consider the probability distribution function of 1D end-to-end distance

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

The probability that the two ends will be within some small distance δ with $\delta \ll \sqrt{N}a$ is

$$e^{-\frac{R^2}{2Na^2}} \approx 1 \quad for -\delta < R < \delta$$

What is this end to end distance concept in a loop. Does it mean that the two ends are almost at negligible distance from each other?

and therefore

$$P(R;N) \approx \left(\frac{1}{2\pi N a^2}\right)^{\frac{1}{2}}$$

where N is the number of Kuhn segments. We integrate over all distances of near contact (1D)

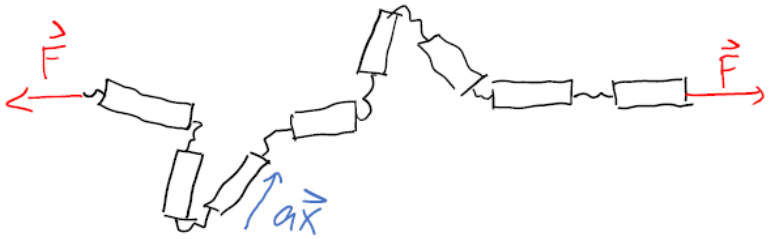
$$p_0^{1D} = \int_{-\delta}^{\delta} \left(\frac{1}{2\pi N a^2}\right)^{\frac{1}{2}} dR = \sqrt{\frac{2}{\pi N}}\frac{\delta}{a}$$

It is important to note that the probability for loop formation decays much more strongly with the chain length in three dimensions (3D):

$$p_0^{3D} = \sqrt{\frac{6}{\pi N^3}}\left(\frac{\delta}{a}\right)^3$$

2.1.6. Force-extension curves from Freely-Jointed Chain model

We will use the FJC model to describe the force required to stretch the polymer to a given end-to-end distance. For free polymer chains, $P(R;N)$ is sharply peaked around R = 0. As the chain is stretched to lengths approaching its contour length, the overall number of conformations goes down, and with it so too does the entropy. This reduction in entropy corresponds to an increase in the free energy.



Remember that the orientation of Kuhn segments is independent from the orientation of its neighbors. This system is equivalent to a succession of paramagnetic spins under the influence of an external field [6]. The energy of the chain can be written as

$$E_{FJC} = -\sum_{i=1}^{N} a\,\vec{F}\vec{x}_i$$

where a is the Kuhn segment length, $\vec{F}$ is the force applied between both ends of the polymer, and $\vec{x}_i$ denotes the orientation of the *ith* Kuhn segment. The partition function is

$$Z = \sum e^{-\frac{E_{FJC}}{kT}}$$

Each link fluctuates independently, therefore the partition function of *N* links is

$$Z_N = Z_1^N$$

and

$$Z_1 = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta\, sin\theta\; e^{\frac{Fa\,cos\theta}{kT}} = 4\pi\frac{kT}{Fa}sinh\left(\frac{Fa}{kT}\right)$$

The free energy is

$$G = -kTlnZ_N$$

The average extension (or end-to-end distance) $\langle z \rangle$ is given by

$$\langle z \rangle = -\frac{\partial G}{\partial F} = Na\left(coth\left(\frac{Fa}{kT}\right) - \frac{kT}{Fa}\right)$$

In the small force limit where $Fa \ll kT$ we can expand $\coth(x) = \frac{1}{x} + \frac{x}{3} + \cdots$ and find

$$\langle z \rangle \approx \frac{Na\,Fa}{3kT} = \frac{L_{tot}Fa}{3kT}$$

or

$$F = k_{spring}\langle z \rangle$$

with $k_{spring} = \frac{3kT}{L_{tot}a}$ (3D). In other words, for small pulling forces, the entropic restoring force of a polymer behaves like a Hookian spring whereby the spring constant depends on the Kuhn segment or persistence length.

2.2. The Worm-Like-Chain model

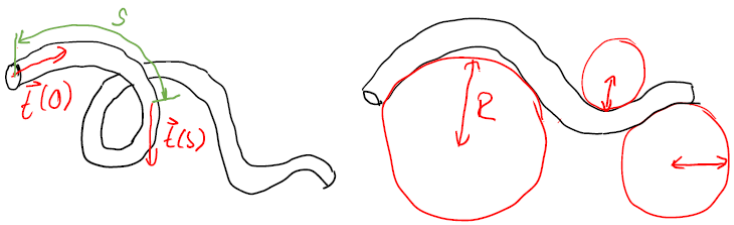
The freely-jointed chain model describes conformations of unperturbed polymers well if the contour length is long compared to the persistence length. The Worm-Like Chain (WLC) model is a more generally applicable model useful for describing stretched or bent DNA molecules as well as cytoskeletal fibers (including actin or microtubules) whose persistence length is comparable to or larger than the size of a cell. To derive the WLC model, we start with the energetics of beam deformation [4].

2.2.1. Beam theory and persistence length

The bending energy of a beam of length L is

$$E_{bend} = \frac{K_{eff}}{2}\int_0^L ds \left|\frac{d\vec{t}}{ds}\right|^2$$

where the flexural rigidity K_{eff} is a material parameter and $\kappa(s) = \frac{1}{R(s)} = \left|\frac{d\vec{t}}{ds}\right|$ is the curvature.



The bending energy of a circular loop with radius R is

$$E_{bend} = \frac{\pi K_{eff}}{R}$$

The persistence length is the length over which a polymer is roughly rigid. Competition between thermal fluctuations and energetic cost associated with beam bending leads to the definition of the persistence length ξ. When elastic energies compete with thermal fluctuations, the relevant comparison is

$$kT \approx \frac{K_{eff} L}{2R^2}$$

where we assume constant curvature over the length of the polymer. Roughly speaking, the persistence length is that length of polymer for which the radius of curvature is equal to the length of polymer itself. Setting L and R to ξ we estimate

$$\xi \approx \frac{K_{eff}}{2kT}$$

Next we will derive the relation between persistence length and flexural rigidity [4]. The persistence length characterizes the correlations in the tangent vectors of different positions along the polymer with an exponentially decaying the tangent-tangent correlation function

$$g(s) = \langle \vec{t}(s)\vec{t}(0) \rangle = e^{-\frac{s}{\xi}}$$

and the limits $g(0) = 1$ and $g(s \to \infty) = 0$. We take a short beam with $s << \xi$. Thermal fluctuations can bend it only slightly and the shape can be approximated by an arc of radius *R*.

$E_{bend} = \frac{K_{eff} s}{2R^2} = \frac{K_{eff}}{2s}\theta^2$ and $g(s) = \langle cos\theta(s) \rangle$. The cosine function can be expanded into a Taylor series for $\theta(s) \ll 1$, i.e. $g(s) = \left\langle 1 - \frac{\theta^2}{2} \right\rangle$.

The thermal average is computed by summing over all possible orientations of the tangent vector at *s*

$$\langle \theta^2(s) \rangle = \frac{1}{Z} \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, sin\theta \, \theta^2 exp\left[-\frac{K_{eff}}{2kTs} \theta^2 \right]$$

where *Z* is the partition function

$$Z = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, sin\theta \, exp\left[-\frac{K_{eff}}{2kTs} \theta^2 \right]$$

or

$$\langle \theta^2(s) \rangle = -2kTs \, \frac{\partial \, ln \, Z}{\partial K_{eff}}$$

For small angles $sin\theta \approx \theta$. We substitute $u = \frac{K_{eff}}{2kTs}\theta^2$

$$Z = \frac{2\pi kTs}{K_{eff}} \int_0^\infty du\, e^{-u} = \frac{2\pi kTs}{K_{eff}}$$

Note that for $s \ll \xi$ the upper bound tends to infinity. Using $\frac{\partial\, lnZ}{\partial K_{eff}} = -\frac{1}{K_{eff}}$ we find

$$g(s) = \left\langle 1 - \frac{\theta^2}{2} \right\rangle = 1 - \frac{kTs}{K_{eff}}$$

With $g(s) = e^{-\frac{s}{\xi}} \approx 1 - \frac{s}{\xi}$ for for $s \ll \xi$

$$\xi = \frac{K_{eff}}{kT}$$

This agrees within a factor of 2 with our estimate.

2.2.2. Elasticity and Entropy: The Worm-Like Chain model

The WLC model accounts for both the elastic energy and entropy of polymer chains [4, 7]. It uses an approximation and interpolation approach to describe the end-to-end distance $\langle z \rangle$ of a polymer under external force. The entropic cost of stretching is

$$E_{FJC} = -\int_0^L \vec{F}\vec{t}(s)ds$$

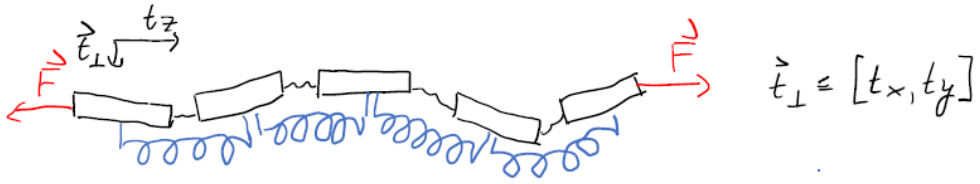
Where do we get these equations from?

The energy cost of bending is

$$E_{bend} = \frac{kT\xi}{2} \int_0^L \vec{\kappa}(s)^2 ds$$

They sum up to

$$E_{WLC} = \frac{kT\xi}{2} \int_0^L \vec{\kappa}(s)^2 ds - \int_0^L \vec{F}\vec{t}(s)ds$$



In the low force regime $F\xi << kT$:

$$F \approx \frac{3kT\langle z \rangle}{4N\xi^2} = \frac{3}{2}\frac{kT\langle z \rangle}{\xi L}$$

as described in 2.1. In the high force regime $F\xi >> kT$, the tangent vectors fluctuate only slightly around the direction of force applied in z-direction. Since the tangent vector is normalized $|\hat{t}| = 1$ we can write $t_z = \sqrt{1 - \vec{t}_\perp^{\,2}}$. Taylor expansion to 2nd order around $t_\perp = 0$

yields $t_z \approx 1 - \frac{1}{2}\vec{t}_\perp^2$ and $\kappa^2 \approx \left(\partial_s \vec{t}_\perp\right)^2$. The energy of the WLC can therefore be approximated by

$$E_{WLC} = \frac{kT}{2}\int_0^L ds \left[\xi\left(\partial_s \vec{t}_\perp\right)^2 + \frac{F}{kT}\vec{t}_\perp^{\,2}\right] - FL$$

Fourier transform decouples energy into normal modes

$$\tilde{t}_\perp(q) = \int ds\, e^{iqs}\, \vec{t}_\perp(s)$$

In Fourier space, the energy takes the form of the potential energy of a collection of harmonic oscillators

$$\frac{E}{kT} = \frac{1}{2}\int \frac{dq}{2\pi}\left[\xi q^2 + \frac{F}{kT}\right]|\tilde{t}_\perp|^2 \; - \frac{F}{kT}L$$

and we can compute the average $\langle t_\perp^2 \rangle$. With the equipartition theorem

$$\left\langle \frac{kT}{2}\left[\xi q^2 + \frac{F}{kT}\right]|\tilde{t}_\perp|^2 \right\rangle = kT$$

we obtain

$$\int \frac{dq}{2\pi}\langle|\tilde{t}_\perp|^2\rangle = 2\int \frac{dq}{2\pi}\frac{1}{\xi q^2 + F/kT} = \frac{1}{(F\xi/kT)^{\frac{1}{2}}}$$

Employing Parseval's theorem $\langle t_\perp^2 \rangle = \int \frac{dq}{2\pi}\langle|\tilde{t}_\perp|^2\rangle$ we obtain the average end-to-end distance

$$\frac{\langle z \rangle}{L} = \langle t_z \rangle \approx 1 - \frac{1}{2}\langle \vec{t}_\perp^2 \rangle = 1 - \frac{1}{2}\frac{1}{(F\xi/kT)^{\frac{1}{2}}}$$

In the WLC model, this result is combined with the low force limit and interpolated for intermediate force

$$F = \frac{kT}{\xi}\left[\frac{1}{4\left(1 - \frac{z}{L}\right)^2} + \frac{z}{L} - \frac{1}{4}\right]$$

The WLC model is in excellent agreement with single-molecule experiments that measure the force required to stretch DNA to a give end-to-end distance.

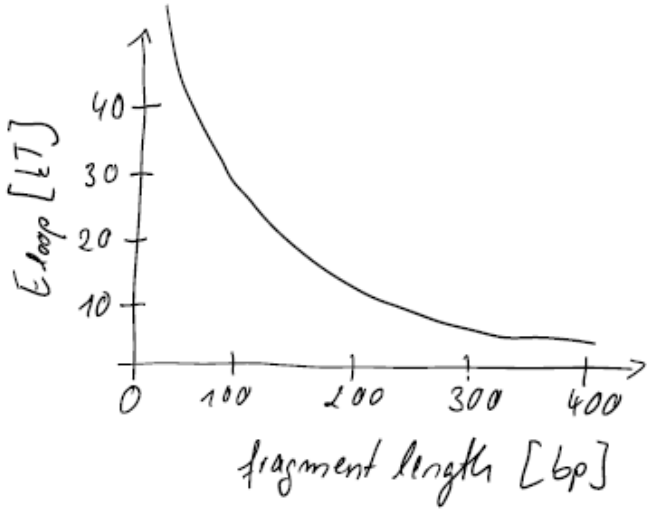
2.2.3. DNA looping redux

In 2.1.5, we have looked into the entropic cost to loop formation; the cost increases as a function of loop length. Here, we will refine the model considering the cost of bending. For a closed loop with radius *R*, the energy is

$$E_{loop} = \frac{\xi \pi kT}{R}$$

The cost of bending decreases with increasing loop length. Considering DNA with a persistence length of ξ = 50 nm and the length of a single baser pair (δ = 0.34 nm) we have

$$\frac{E_{loop}}{kT} = \frac{2\pi^2}{N_{bp}} \frac{\xi}{\delta} \approx \frac{3000}{N_{bp}}$$

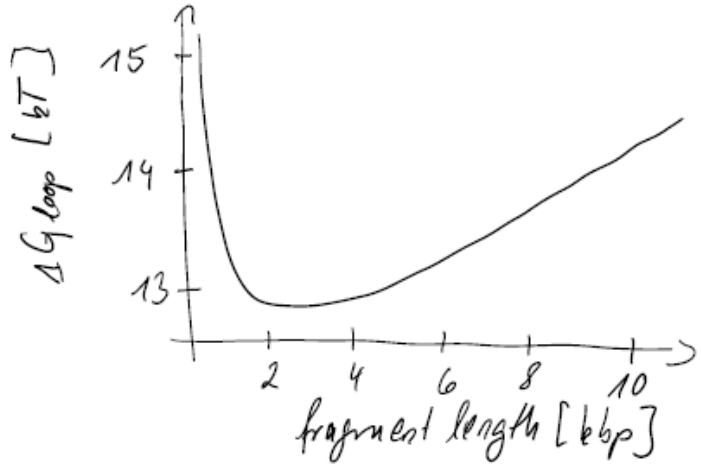


Taking both contributions into account, we calculate the free energy change upon loop formation as a function of loop length. The entropy loss due to loop formation is

$$\Delta S_{loop} = k \, ln p_0 = k \left(-\frac{3}{2} ln N_{bp} + const \right)$$

and thus the free energy change

$$\Delta G_{loop} = \Delta E_{loop} - T \Delta S_{loop} \approx kT \left(\frac{3000}{N_{bp}} + \frac{3}{2} ln N_{bp} + const \right)$$

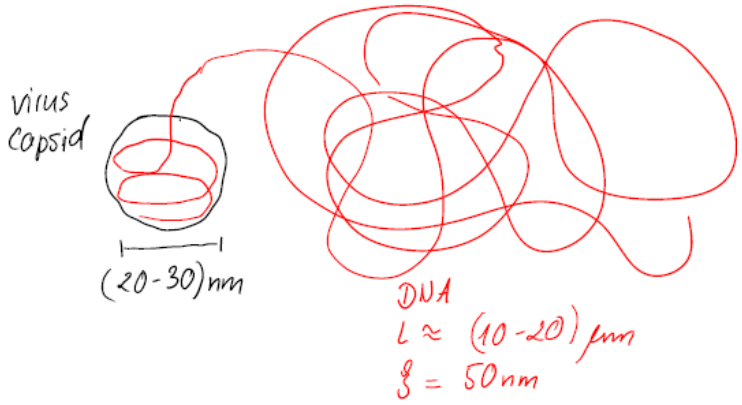


Experimentally, the probability of loop formation as a function of loop length has been addressed [4]. The idea was to trap spontaneous thermal fluctuations of end-to-end distance

when the ends are in close proximity. To this end, an enzyme called ligase was added to DNA molecules with different length. The ligase forms a bond between both ends of the DNA when they are in close proximity. The probability of loop formation is directly proportional to the fraction of ligated loops.

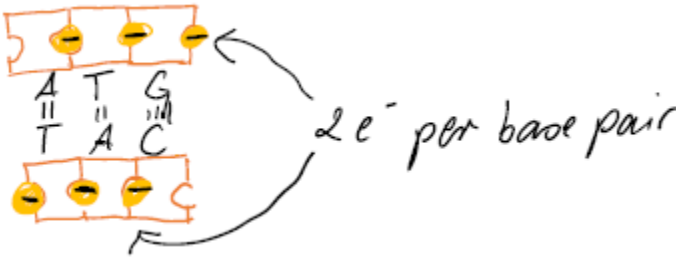
2.3. DNA packaging and the electrostatics of salty solutions

In cells and viruses, DNA is tightly packed. Estimate the typical length of genomic DNA of a virus, a bacterium, and a eukaryotic cell and compare to the diameters of their containers.



The radius of gyration of genomic DNA molecules exceeds the size of their containers. The packaging ratio $\nu = \frac{size\ of\ genome}{size\ of\ container} \approx \frac{N_{bp} \cdot 1nm^3}{\frac{4}{3}\pi R^3}$ is ν(virus) ≈ 0.6 for viruses and ν(bacterium) ≈ 0.1. Even more dramatically, the diameter of a typical virus capsid is smaller than the persistence length of DNA. We can therefore expect an entropic cost of packaging and a cost of bending. Before we consider these two costs, however, we will focus on the electrostatics of DNA packaging.

2.3.1. Electrostatic energy of a charged ball



DNA is strongly negatively charged. It carries one e^- per phosphate group or 2 e^- per base pair. We start estimating the free energy of DNA packaged into a virus capsid by calculating the electrical energy of a ball of radius *R* and charge *Q* assuming that the charge is homogeneously

distributed throughout the ball. The electrical potential of a charged ball at position *r* is given by

$$V(r) = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q}{r}$$

and the charge density is $\rho = \frac{Q}{4/3\pi R^3}$. The work performed for assembling the ball is calculated by adding shells with charge $dq = \rho 4\pi r^2 dr$. When adding layers to the onion, adding the layer at radius *r* costs an energy dictated by the potential of the charge already there *V(r)* times the charge of the added shell.

$$dU_{el} = V(r)dq = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{\rho\frac{4}{3}\pi r^3}{r}\rho 4\pi r^2 dr$$

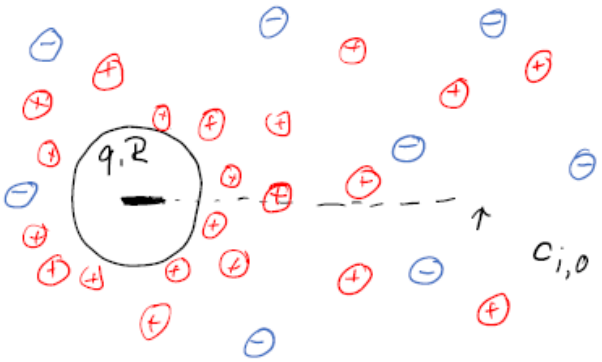
$$U_{el} = \int_0^R \frac{4\pi}{3\varepsilon_0\varepsilon_r}\rho^2 r^4\, dr = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{3}{5}\frac{Q^2}{R}$$

We can estimate the work required for packaging DNA into a virus. Experimentally, this work has been determined for a bacterial virus called bacteriophage Φ29 [8]. The radius is $R_{\Phi 29} \approx 20$ nm and the length is $N_{\Phi 29} \approx 2{\cdot}10^4$ bp. Thus $U_{el} \approx 10^8 pN\ nm$. Smith et al used laser tweezers for determining the work required for packaging DNA into the bacteriophage and found $W_{interalization} \approx 2 \cdot 10^5\ pN\ nm$. How can the discrepancy of three orders of magnitude be explained?

2.3.2. The Poisson-Boltzmann equation

The charge carried by DNA (and other biological polymers) is screened by the condensation of counter-ions. Biological cells carry different ion species at a typical concentration of 150 mM. The energy necessary to liberate ions from molecules can be comparable to the thermal energy. Two equal opposite charges are separated by a distance $l_B = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r kT} \approx 0.7\ nm$ (Bjerrum length) such that their electrostatic interaction energy is equal to their thermal energy.

Salt dissociates in water and ions move freely. As a consequence, a cloud of counter-ions forms around a charged polymer. The distance over which the cloud extends is determined by a competition between energy and entropy of the ions. In the following, we aim at finding the characteristic length over which the cloud of counter-ions decays. Consider different ions *i* with charge z_i as point-like charges.



The solution is neutral and therefore $\sum_{i=1}^{n} c_{i,0} z_i = 0$, where $c_{i,0}$ is the concentration of ions far away from the polymer. Consider the charge density in the vicinity of the charged polymer with charge q and radius R. The Poisson equation provides the relation between the electrostatic potential V and the charge density ρ

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0 \varepsilon_r}$$

(V depends on the central charge and the charge of the counter-ions. The contribution of the central charge enters through the integration constant a; see below.) The concentration of counter-ions c_i follows the Boltzmann distribution

$$c_i(r) = c_{i,0}\, exp\left(-\frac{z_i e V(r)}{kT}\right)$$

whereby the charge density relates to the concentration as $\rho(r) = \sum_{i=1}^{n} z_i e c_i(r)$. z_i is the valence of ion species i. Taken together, we find the Poisson-Boltzmann equation

$$\nabla^2 V = -\sum_{i=1}^{n} \frac{z_i e c_{i,0}}{\varepsilon_0 \varepsilon_r}\, exp\left(-\frac{z_i e V(r)}{kT}\right)$$

This equation can be solved analytically in the limit eV << kT (Debye-Hückel approximation). Expansion to second order yields

$$\nabla^2 V(r) = \sum_{i=1}^{n} \frac{z_i e c_{i,0}}{\varepsilon_0 \varepsilon_r} + \frac{z_i^2 e^2 c_{i,0}}{\varepsilon_0 \varepsilon_r kT} V(r) + \cdots$$

The first term disappears due to charge neutrality far away from the charged polymer. The Debye length is defined as $l_D^{-1} = \sqrt{\sum_{i=1}^{n} \frac{z_i^2 e^2 c_{i,0}}{\varepsilon_0 \varepsilon_r kT}}$. Using spherical symmetry $\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V(r)}{\partial r}\right) = \frac{1}{l_D^2} V(r)$ we find the solution

$$V(r) = \frac{a}{r} exp\left(-\frac{r}{l_D}\right)$$

To determine the integration constant a, we calculate the electric field at the surface of the polymer with radius R using Gauss' theorem $\oint d\vec{A}\vec{n} = \int dV\, \nabla\vec{E}$ and Gauss' law $\nabla\vec{E} = \frac{\rho}{\varepsilon_0\varepsilon_r}$ and, therefore, $4\pi R^2 E(r) = \frac{q}{\varepsilon_0\varepsilon_r}$

$$E = -\nabla V = a\frac{exp\left(-\frac{r}{l_D}\right)}{r}\left(l_D^{-1} + r^{-1}\right)$$

$$a = \frac{q}{4\pi\varepsilon_0\varepsilon_r}\frac{exp\left(\frac{R}{l_D}\right)}{1+\frac{R}{l_D}}$$

and

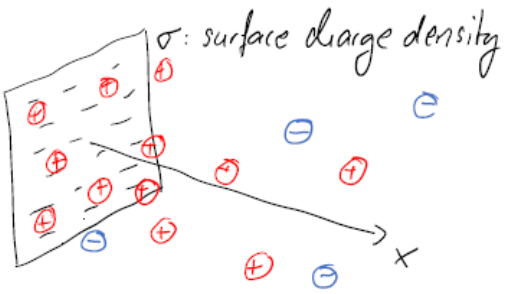
$$V(r) = \frac{q}{4\pi\varepsilon_0\varepsilon_r}\frac{exp\left(\frac{R}{l_D}\right)}{1+\frac{R}{l_D}}\frac{exp\left(-\frac{r}{l_D}\right)}{r}$$

We find that the electrostatic potential of the charged polymer is strongly screened by the cloud of counter-ions. By multiplying with the charge ($Q = e$) of a second ion we can estimate the electrostatic interaction energies in different media. $q = e$ and $R = 0.5$ *nm*.

Distance [nm]	**Vaccum** $\varepsilon_r = 1$ $l_D^{-1} = 0$ nm^{-1}	**De-ionized water** $\varepsilon_r = 80$ $l_D^{-1} = 0$ nm^{-1}	**Physiological salty solution** $\varepsilon_r = 80$ $l_D^{-1} = 1.27$ nm^{-1}
0.1	14.4 eV 576 kT	0.180 eV 7.2 kT	0.18 eV 7.2 kT
1	1.44 eV 57.6 kT	0.018 eV 0.72 kT	$5.11\cdot10^{-3}$ eV 0.204 kT
10	0.144 eV 5.76 kT	0.002 eV 0.072 kT	$5.77\cdot10^{-9}$ eV $2.308\ 10^{-7}$ kT

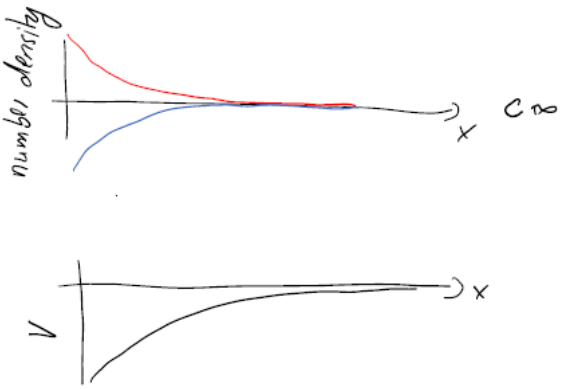
Close to a charged surface, the electrostatic potential decays like

$$V(x) = \frac{\sigma l_D}{\varepsilon_0\varepsilon_r}\, exp\left(-\frac{x}{l_D}\right)$$



Using the Poisson equation, we can write down how the density of counter-ions decays as a function of the distance *x* from the charged surface

$$\rho(x) = \frac{\sigma}{l_D}\, exp\left(-\frac{x}{l_D}\right)$$



2.3.3. DNA packaging into a virus capsid

With the idea of screening though counter-ions at hand, we use a different approach for estimating the work performed while packaging DNA into the virus capsid. We assume that the charge of DNA is fully screened through positive counter-ions which enter the capsid with the DNA and we neglect negatively charged ions. The free energy cost of packaging is then the result of decreasing the entropy of the charges that make up the cloud. The derivation of the free energy of dilute solutions may be found in chapter 6.2.2 of [4]. With

$$p\frac{dV}{dN} = \frac{dG}{dN} = kTln\left(\frac{c_{capsid}}{c_{cloud}}\right)$$

the work for squeezing the counter-ions into the capsid is

$$W_{el} = NkTln\left(\frac{V_{cloud}}{V_{capsid}}\right)$$

With $c_{i,0} = 100mM$ the Debye length is $l_D = \sqrt{\frac{1}{4\pi l_b c_{i,0}}} \approx 1nm.$ Thus $V_{cloud} = L\pi((R_{DNA} + l_D)^2 - R_{DNA}^2) \approx 64000\, nm^3$ and $V_{capsid} = \frac{4\pi}{3}R_{capsid}^3 - L\pi R_{DNA}^2 \approx 12000\, nm^3$ we find $W_{el} \approx 65000$ kT = $2.6 \cdot 10^5\, pN\, nm$ in good agreement with experiments [4, 8].

More sophisticated models allow determining the force that resists packaging form the free energy of packaged DNA

$$G_{tot} = G_{charge} + G_{bend} + G_{entropic}$$

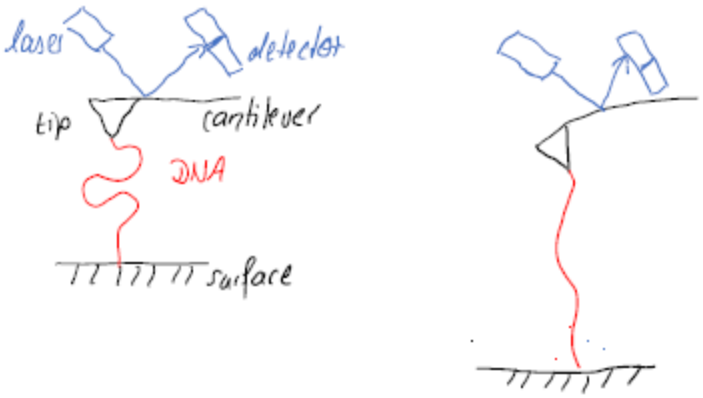
Next to the charge term, the bending energy and the entropic cost of confining DNA are taken into account [4].

2.4. Biophysical characterization of DNA elasticity

Various biophysical techniques have been developed for characterizing elastic properties of single DNA molecules, studying the kinetics of DNA-binding enzymes and molecular motors, and characterizing DNA-uptake by viruses and bacteria. All of them allow manipulating single-molecules.

2.4.1. Atomic force microscopy

The most widely used technique is atomic force microscopy (AFM). AFM was originally developed for scanning surfaces, generating topographical images. Under certain conditions they allow acquiring movies of conformational changes in single molecules [9]. In addition to imaging, AFM can be used for force spectroscopy. The molecule of interest (e.g. a DNA molecule) is attached to a surface at one end and to the tip of the AFM at the other end. As the surface is moved away from the tip, force is exerted onto the DNA molecule. As a consequence, the cantilever - to which the tip is attached - bends. A laser beam is used as a readout of cantilever bending. If the elastic properties of the cantilever are calibrated experimentally, the degree of bending can be converted into the force acting on the DNA. The distance between the tip and the surface provides the end-to-end distance of the DNA molecule. This assay allows determining the force-extension relation (2.2.2) experimentally. AFMs can generate large forces in the range of nanonewtons but they are limited in resolution. Better resolution at forces below 5 pN can be achieved using laser tweezers or magnetic tweezers [10].



2.4.2. Laser tweezers

Laser tweezers or optical traps are versatile tools for studying molecular processes at the single molecule level. The Nobel Prize in physics was awarded to their inventor, Arthur Ashkin, in 2018. To generate a laser trap, laser light is focused to a diffraction-limited spot by the objective of a light microscope. As a consequence, a strong light gradient is generated near the focus. When a dielectric object resides in proximity of the laser focus, force is generated. Laser traps used in biophysics usually work in aqueous environment and, therefore, the refractive index of the trapped particle (e.g. a radius $r \approx 1$ µm latex bead) must be higher compared to water to

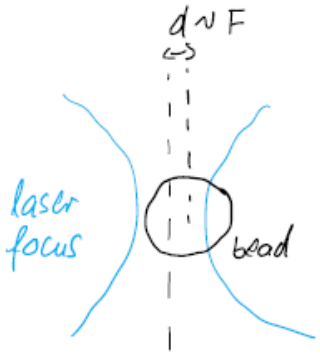
allow for trapping. Force generation by laser traps can be theoretically understood in two regimes $r >> \lambda$ and $r << \lambda$, respectively. In the second limit, the force acting on the particle, or gradient force can be derived calculating the Lorentz force on a dipole in an electromagnetic field. The Lorentz force on a point dipole with dipole moment $\vec{p} = \alpha\vec{E}$ is

$$\vec{F} = (\vec{p}\nabla)\vec{E} + \frac{1}{c}\frac{\partial\vec{p}}{\partial t} \times \vec{B}$$

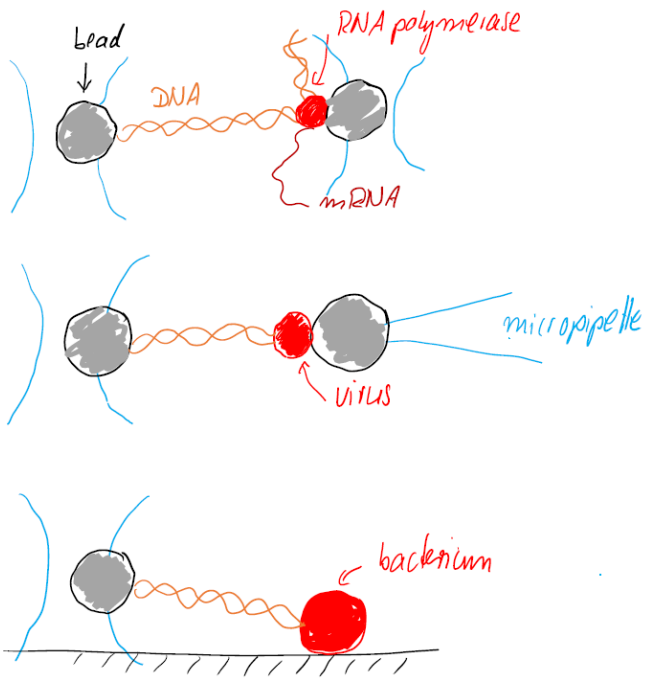
where α is the polarizability. Using Maxwell's equations and considering only continuous power lasers we find that the force is proportional to the gradient of the light intensity $\nabla I = \nabla\vec{E}^2$

$$\vec{F} = \frac{1}{2}\alpha\nabla\vec{E}^2$$

Close to the laser focus, the trap potential is quadratic and therefore the optical restoring force increases linearly as a function of the deflection d of a particle from the center of the laser trap with $F = k_{trap}d$. The trap stiffness k_{trap} has to be calibrated experimentally.

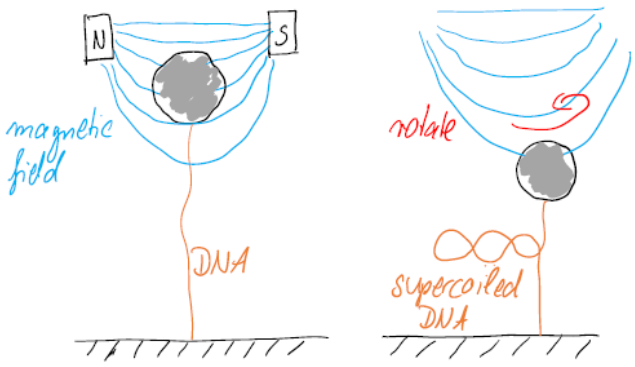


Using laser tweezers, forces as small as 1 pN can be measured and thus they were employed for verifying the WLC model [4]. Importantly, laser tweezers can be used to detect positional changes with an accuracy of 0.1 nm, the diameter of a hydrogen atom [10]. It has been used to measure the step size of 0.34 nm of RNA polymerases, the enzymes that catalyze the synthesis of messenger RNA [11]. Laser tweezers were also used to measure the work performed during DNA packaging into a bacterial virus [8] and to characterize the kinetics of bacterial DNA uptake [12].



2.4.3. Magnetic tweezers

High force resolution down to 0.05 pN can be achieved using magnetic tweezers [10] (at the cost of low spatial and temporal resolution). Superparamagnetic beads are attached to one end of a DNA molecule and the other end of DNA binds to a surface. Magnetic fields then induce a magnetization $\vec{M}$ and thus a torque $\vec{T} = \vec{M} \times \vec{B}$. The force pulling the bead upwards in the direction of the field gradient is given by $\vec{F} \propto \nabla \vec{B}^2$. Again the stiffness of the trap has to be calibrated experimentally [10]. Importantly, magnetic tweezers can rotate the bead and, therefore, apply torque onto the DNA molecule with leads to supercoiling. DNA supercoiling is highly relevant in biological cells. Given the double-helical nature of DNA, the state of supercoiling has to be controlled by dedicated enzymes, so-called topoisomerases [1]. When DNA replicates, the double-helix must open up. Since bacterial chromosomes are circular, pulling both strands of the DNA apart would generate supercoiling and this would ultimately inhibit replication. Using magnetic tweezers, the kinetics of topoisomerases have been studied [13].



1. Alberts, B., et al., *Molecular Biology of the Cell*. 2015, Garland Science.
2. Rosenfeld, N., et al., *A fluctuation method to quantify in vivo fluorescence data.* Biophysical Journal, 2006. **91**(2): p. 759-766.
3. Rosenfeld, N., et al., *Gene regulation at the single-cell level.* Science, 2005. **307**(5717): p. 1962-1965.
4. Phillips, R., et al., *Physical Biology of the Cell*. 2013: Garland Science.
5. Golding, I., et al., *Real-time kinetics of gene activity in individual bacteria.* Cell, 2005. **123**(6): p. 1025-1036.
6. Grosberg, A.Y. and A.R. Khokhlov, *Statistical Physics of Macromolecules*. 1994, Woodbury, NJ: AIP press.
7. Marko, J.F. and E.D. Siggia, *Stretching DNA.* Macromolecules, 1995. **28**(26): p. 8759-8770.
8. Smith, D.E., et al., *The bacteriophage phi 29 portal motor can package DNA against a large internal force.* Nature, 2001. **413**(6857): p. 748-752.
9. Ando, T., T. Uchihashi, and N. Kodera, *High-speed AFM and applications to biomolecular systems.* Annu Rev Biophys, 2013. **42**: p. 393-414.
10. Greenleaf, W.J., M.T. Woodside, and S.M. Block, *High-resolution, single-molecule measurements of biomolecular motion.* Annual Review of Biophysics and Biomolecular Structure, 2007. **36**: p. 171-190.
11. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase.* Nature, 2005. **438**(7067): p. 460-5.
12. Hepp, C. and B. Maier, *Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism.* Proc Natl Acad Sci U S A, 2016. **113**(44): p. 12467-12472.
13. Strick, T.R., V. Croquette, and D. Bensimon, *Single-molecule analysis of DNA uncoiling by a type II topoisomerase.* Nature, 2000. **404**(6780): p. 901-904.
14. Sackmann, E. and R. Merkel, *Lehrbuch der Biophysik*. 2010, Weinheim: Wiley-VCH.